

Klassifizierung von Pollenproben mit spektroskopischen und spektrometrischen Methoden und multivariater Statistik

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m

(Dr. rer. nat.)

im Fach Chemie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

Dipl.-Chem. Stephan Seifert

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät

Prof. Dr. Elmar Kulke

Gutachter: 1. Prof. Janina Kneipp
 2. Prof. Achim Kohler

Eingereicht am: 26.01.2016

Tag der mündlichen Prüfung: 25.04.2016

Kurzzusammenfassung

In dieser Arbeit werden spektroskopische und spektrometrische Methoden diskutiert, um Verfahren auf Grundlage von umfassenden, molekularen Informationen für die Charakterisierung, Klassifizierung und Identifizierung von Pollen zu entwickeln. Es ist bereits bekannt, dass Raman- und Infrarot-Spektroskopie spezifische Fingerabdrücke der chemischen Zusammensetzung von Pollen liefern können. Im Rahmen dieser Arbeit werden nun die oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS), die *matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS) und die Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie) einzelner Pollenkörner für die taxonomische Klassifizierung eingesetzt.

Die Klassifizierung mit SERS basiert auf Unterschieden der inneren, zellulären Bestandteile und bietet verglichen mit normaler Raman-Spektroskopie mehrere Vorteile für die Identifizierung von Pollen. Es wird gezeigt, dass künstliche neuronale Netze für die Klassifizierung und Vorselektion von Spektren eingesetzt werden können.

MALDI-TOF MS liefert Fingerabdrücke, die auf Glykoproteine der Pollenkörner zurückgeführt werden können. Da es sich bei vielen der allergenen Bestandteile von Pollen um Glykoproteine handelt, ist die in diesem Zusammenhang vorgestellte Klassifizierungsmethode von besonderem Interesse.

Die taxonomische Klassifizierung von FTIR-Spektren einzelner Pollenkörner wurde bisher durch die Überlagerung der reinen Absorptionsspektren durch Streuartefakte verhindert. In dieser Arbeit werden mathematische Methoden eingesetzt, um dieses Problem zu überwinden. Da Pollenkörner komplexe, biologische Partikel sind, die aus vielen verschiedenen Biomolekülen bestehen, ist es sinnvoll, für ihre Untersuchung komplementäre Methoden zu verknüpfen. Dies wird durch eine *Consensus Principal Component Analysis* (CPCA) erreicht. Neben der Anwendung zur taxonomischen Klassifizierung können spektroskopische und spektrometrische Methoden auch zur Erkennung von Unterschieden, die durch Umwelteinflüsse verursacht werden, eingesetzt werden. Als ein erstes Beispiel dieser Anwendung werden *cherry leaf roll virus* (CLRV) infizierte Birkenpollen untersucht und mit Pollen von Kontrollpflanzen verglichen.

Die Ergebnisse dieser Arbeit zeigen, dass spektroskopische und spektrometrische Methoden vielversprechend für die Pollenanalytik in vielen verschiedenen Forschungsgebieten sind.

Abstract

The investigation of pollen grain samples for characterization, classification and identification is still an analytical task that is both challenging and time-consuming, since it is mainly based on the morphological characteristics of the pollen grains. In order to develop approaches that are based on extensive molecular information and that lead to an automated classification, in this thesis, a combination of spectroscopic and spectrometric tools is discussed. It is already known that a fingerprint analysis of the chemical composition of pollen samples can be accomplished by Raman and infrared spectroscopies. Here, surface-enhanced Raman scattering (SERS), matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS), and Fourier transform infrared (FTIR) spectroscopy of single pollen grains are successfully applied for the taxonomic classification of pollen samples.

It is demonstrated that SERS provides several advantages compared to normal Raman scattering for pollen identification and reveals species-specific differences in the pollen phenome that display in the pollen cells. Specifically, classification and pre-selection of SERS spectra with artificial neural networks (ANN) are shown.

MALDI-TOF MS of pollen samples yields fingerprints that can be assigned to glycoproteins. This is of crucial interest, since glycoproteins make up the majority of allergenic substances in pollen. A classification approach based on this molecular information is introduced.

FTIR spectroscopy of single pollen grains for taxonomic classification has not been possible so far, since the pure absorption spectra are superimposed by scattering-based artifacts that change the characteristic fingerprints. Here, it is shown that this problem can be overcome using mathematical tools for data processing.

Since pollen grains are highly complex biological particles that consist of many different biomolecules, the data from the different, complementary approaches should be combined. Classification, exploiting as much molecular information as possible is accomplished by the application of the multiblock method Consensus Principal Component Analysis (CPCA).

In addition to the application for taxonomic classification, it is conceivable that spectroscopic and spectrometric data could also be used to investigate chemical changes caused by environmental influences. As a first example for such an application, cherry leaf roll virus (CLRV) infected birch pollen are investigated and compared to samples from control plants.

The results in this work indicate that spectroscopic and spectrometric methods are powerful analytical tools that may be useful for improved pollen investigation in different areas of research.

Danksagung

Mein ganz besonderer Dank gilt meiner Betreuerin Prof. Janina Kneipp, die mir die Möglichkeit gab, dieses interessante Thema nicht nur mit dem nötigen Freiraum zur Entwicklung und Verwirklichung eigener Ideen zu bearbeiten, sondern mir stets auch durch hilfreiche und inspirierende Diskussionen zur Seite stand.

Weiterhin möchte ich Prof. Achim Kohler, Dr. Boris Zimmermann und Murat Bagcioglu für die anregenden Kooperationstreffen in Deutschland und Norwegen danken. Diese haben meine Promotionszeit in vielerlei Hinsicht bereichert und einen wesentlichen Beitrag zum Gelingen dieser Arbeit beigetragen.

Prof. Ulrich Panne spreche ich meinen Dank dafür aus, dass ich einen Großteil meiner Promotion an der Bundesanstalt für Materialforschung und –prüfung durchführen konnte.

Für die fruchtbare Zusammenarbeit bei MALDI-TOF MS Untersuchungen möchte ich meinen Dank an Dr. Steffen Weidner ausdrücken. Mein Dank gilt in diesem Zusammenhang auch Benjamin Krause, der zahlreiche MALDI-TOF MS-Spektren aufgenommen hat.

Meinen Bürokollegen Maike Joester, Lisa Batzdorf und Nicolai Zientek bin ich für die freundschaftliche Atmosphäre und die interessanten Diskussionen sehr dankbar.

Vielen Dank auch an Dr. Ulrich Schade und Dr. Eglof Ritter ohne deren tatkräftige Unterstützung die Aufnahme von Synchrotron-FTIR-Spektren nicht möglich gewesen wäre.

Anka Kohl bin ich für ihre Hilfsbereitschaft bei der Untersuchung von Kaliumbromidpresslingen mit FTIR-Spektroskopie zu großem Dank verpflichtet.

Für die Hilfe beim Sammeln von Pollenproben möchte ich meinen Dank an Thomas Dürbye und Carsten Später ausdrücken.

Auch möchte ich mich bei Prof. Rademann und Christian Schaumberg für die gute Kooperation bei den ESEM-Untersuchungen bedanken.

Prof. Carmen Büttner und Dr. Maria Landgraf möchte ich für die Möglichkeit danken, im Rahmen meiner Arbeit CLRV-infizierte Pollenproben untersuchen zu können.

Vielen Dank auch an die ganze Arbeitsgruppe Kneipp für die gute Zusammenarbeit und die freundschaftliche Stimmung in der Gruppe und besonders an Sabrina Diehn, Christian Torres-Ziegenbein und Marina Gühlke für die Unterstützung bei der SERS-Untersuchung CLRV-infizierter Pollenproben. Dr. Virginia Merk danke ich für das Korrekturlesen dieser Arbeit.

Zu guter Letzt möchte ich meiner Familie und meinen Freunden danken, die mir fortwährend mit ihrer Unterstützung zur Seite standen. Ganz großer Dank gilt meiner Frau Elina für das Beseitigen zahlreicher Fehler in meiner Arbeit, sowie für das entgegengebrachte Verständnis, insbesondere dann, als die Dissertation ständiger (gedanklicher) Begleiter war.

Inhaltsverzeichnis

1. Einleitung	1
2. Überblick über die verwendeten Verfahren zur Klassifizierung von Pollenproben	5
2.1 Analytische Verfahren zur Untersuchung von Pollenproben.....	5
2.1.1 Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie)	7
2.1.2 Raman-Streuung	9
2.1.3 Oberflächenverstärkte Raman-Streuung (<i>surface-enhanced Raman scattering</i> , SERS) .	10
2.1.4 <i>Matrix-assisted laser desorption/ionisation time-of-flight</i> -Massenspektrometrie (MALDI-TOF MS).....	12
2.1.5 Environmental scanning electron microscopy (ESEM).....	13
2.2 Multivariate Verfahren	14
2.2.1 Hierarchische Clusteranalyse (HCA)	14
2.2.2 Hauptkomponentenanalyse (<i>Principal Component Analysis</i> , PCA)	16
2.2.3 <i>Partial least squares regression</i> (PLSR)	17
2.2.4 <i>Consensus Principal Component Analysis</i> (CPCA)	19
2.2.5 Künstliche neuronale Netze (<i>artificial neural networks</i> , ANN)	21
3. Materialien und Methoden	23
3.1 Pollenproben.....	23
3.2 Raman-Apparatur	27
3.3 Raman-Untersuchungen am Raffinat des Pollenextrakts	27
3.4 SERS-Untersuchungen von Pollenextrakten mit Nanopartikeln in Lösung.....	27
3.5 MALDI-TOF MS-Untersuchungen.....	28
3.6 FTIR-Untersuchungen einzelner Pollenkörner.....	29
3.7 FTIR-Untersuchungen von Kaliumbromidpresslingen CLRV-infizierter Birkenpollenproben	30
3.8 ESEM Untersuchungen von Pollenkörnern.....	30
3.9 Vorbehandlung und Auswertung von Pollendaten.....	30

3.10	Spektrenvorbehandlung und Auswertung der SERS-Spektren durch HCA und PCA	30
3.11	Klassifizierung von SERS-Spektren durch künstliche neuronale Netze.....	31
3.12	Extraktion von Spektren aus SERS-Daten von Pollenproben aus dem Feld	33
3.13	Spektrenvorbehandlung und multivariate Auswertung der MALDI-TOF MS-Spektren.....	34
3.14	Klassifizierung von FTIR-Spektren einzelner Pollenkörner.....	34
3.15	Spektrenvorbehandlung und multivariate Auswertung von FTIR-Spektren von Kaliumbromidpresslingen	35
3.16	Spektrenvorbehandlung und multivariate Auswertung von Pollendaten durch CPCA.....	35
4.	SERS zur taxonomischen Klassifizierung von Pollen.....	37
4.1	SERS von Pollenextrakten liefert komplementäre schwingungsspektroskopische Informationen	38
4.1.1	Qualitative Diskussion von SERS-Spektren wässriger Pollenextrakte	38
4.1.2	Raman-Untersuchung des Raffinats	40
4.2	Eignung von SERS in Nanopartikellösung zur Klassifizierung von Pollenproben.....	44
4.3	Klassifizierung von SERS-Daten mit multivariater Verfahren.....	46
4.3.1	HCA von SERS-Spektren wässriger Pollenextrakte.....	49
4.3.2	PCA von SERS-Spektren wässriger Pollenextrakte	51
4.3.3	Unterscheidung verschiedener Pollenproben durch Auswertung mit ANN	52
4.4	Wiedergabe taxonomischer Verwandtschaftsverhältnisse aus SERS-Daten	53
4.5	Klassifizierung vorselektierter Spektren des Datensatzes	56
4.5.1	Vorselektion durch Extraktion mit dem ANN.....	56
4.5.2	Untersuchung taxonomisch spezifischer Daten der Ordnungen.....	57
4.5.3	Untersuchung taxonomisch relevanter Daten der Ordnungen.....	59
4.5.4	Untersuchung taxonomisch spezifischer Daten der Gattungen	60
4.5.5	Untersuchung taxonomisch relevanter Daten der Gattungen	64
4.6	Schlussfolgerungen	68
5.	MALDI-TOF MS zur taxonomischen Klassifizierung von Pollen.....	69
5.1	Eignung von MALDI-TOF MS zur Klassifizierung von Pollenproben	70

5.2	Wiedergabe taxonomischer Verwandtschaftsverhältnisse aus MALDI-TOF MS-Daten	72
5.2.1	HCA von MALDI-TOF MS-Spektren.....	73
5.2.2	PCA von MALDI-TOF MS-Spektren.....	77
5.3	Schlussfolgerungen	84
6.	CPCA von MALDI-TOF MS- und SERS-Daten zur taxonomischen Klassifizierung von Pollen	85
6.1	CPCA zur Klassifizierung auf taxonomischer Ebene der Ordnung	86
6.2	CPCA zur Klassifizierung auf taxonomischer Ebene der Gattung.....	89
6.3	Schlussfolgerungen	94
7.	Klassifizierung von FTIR-Daten einzelner Pollenkörner	95
7.1	Eignung der FTIR-Spektroskopie für die Klassifizierung einzelner Pollenkörner	97
7.2	Wiedergabe taxonomischer Verwandtschaftsverhältnisse aus den FTIR-Daten einzelner Pollenkörner	108
7.3	Schlussfolgerungen	121
8.	Klassifizierung CLRV-infizierter Birkenpollen	123
8.1	Untersuchung von Birkenpollenproben aus dem Jahr 2011	124
8.1.1	Klassifizierung CLRV-infizierter Birkenpollenproben mit SERS	124
8.1.2	Klassifizierung CLRV-infizierter Birkenpollenproben mit MALDI-TOF MS	125
8.1.3	Klassifizierung von Kaliumbromidpresslingen CLRV-infizierter Birkenpollenproben mit FTIR-Spektroskopie	128
8.1.4	Kombination verschiedener analytischer Daten zur Klassifizierung CLRV-infizierter Birkenpollenproben	130
8.2	Untersuchung von Birkenpollenproben aus dem Jahr 2012.....	132
8.2.1	Klassifizierung CLRV-infizierter Birkenpollenproben mit SERS	133
8.2.2	Klassifizierung CLRV infizierter Birkenpollenproben mit MALDI-TOF MS	133
8.2.3	Klassifizierung von Kaliumbromidpresslingen CLRV-infizierter Birkenpollenproben mit FTIR-Spektroskopie	135
8.3	Schlussfolgerungen	138
9.	Zusammenfassung und Ausblick.....	139

10. Anhang.....	143
10.1 Abkürzungsverzeichnis.....	143
10.2 Abbildungsverzeichnis.....	145
10.3 Tabellenverzeichnis.....	156
10.4 Literaturverzeichnis	157
10.5 Publikationsverzeichnis	Fehler! Textmarke nicht definiert.

1. Einleitung

Pollenkörner sind die männlichen Keimzellen von Samenpflanzen. Es ist ihre Aufgabe, den haploiden Chromosomensatz für die Fortpflanzung zum weiblichen Fruchtknoten zu transportieren. Um dies zu erreichen, können Pollen auch extremen Umwelteinflüssen trotzen und sich auf unterschiedlichste Weise fortbewegen, wie zum Beispiel durch Wind, Wasser oder auch durch das Anheften an kleine Tiere. Abbildung 1.1 zeigt schematisch den Aufbau eines Pollenkorns. Prinzipiell kann zwischen der Außenhülle, dem sogenannten Sporoderm, und den inneren, zellulären Bestandteilen unterschieden werden.

Der Sporoderm setzt sich aus der inneren Intine, die hauptsächlich Pektin und Cellulose enthält, und der äußeren Exine zusammen. Die Exine besteht u.a. aus Flavonoiden^[1], Carotinoiden^[2], Lipiden^[3] und dem Biomakromolekül Sporopollenin^[4], dessen Aufbau und Struktur bis heute noch nicht vollständig aufgeklärt wurde. Sie ist sehr resistent gegen unterschiedlichste äußere Einflüsse und ihre Hauptaufgabe besteht darin, das Polleninere vor diesen zu schützen. Auf der Oberfläche der Exine befinden sich charakteristische Strukturelemente, die die taxonomisch spezifische Textur des Pollenkorns maßgeblich beeinflussen, sowie der klebrige Pollenkitt, der neben zahlreichen anderen Funktionen auch für das Anheften der Pollenkörner an Tieren oder anderen Transportmitteln verantwortlich ist.^[5] Die meisten Pollenkörner besitzen Einkerbungen in der Exinschicht, sogenannte Aperturen. Durch diese kann bei der Befruchtung der Pollenschlauch wachsen. Im Inneren der Pollenkörner befinden sich neben den männlichen Keimzellen und Ölkörpern, die als Energiespeicher dienen, typische Zellbestandteile wie zum Beispiel Vesikel und endoplasmatisches Retikulum (ER), aber auch Bestandteile, die als Allergene bezeichnet werden.^[7]

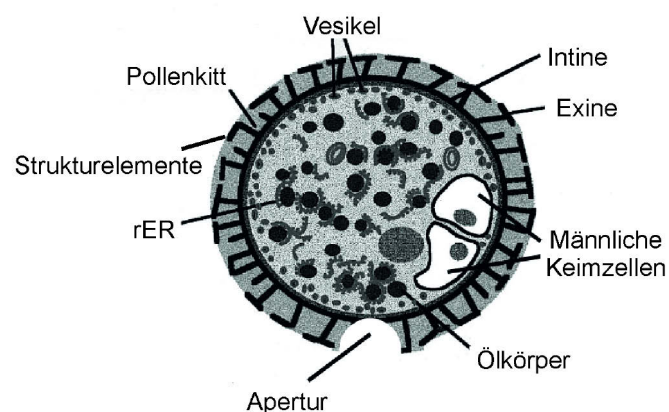


Abbildung 1.1: Schematische Darstellung eines Pollenkorns nach [6].

Pollen sind für viele Forschungsgebiete interessant und werden z.B. in den Agrarwissenschaften, der Paläobiologie, Klimaforschung oder Medizin untersucht. Das Interesse in der Medizin beruht darauf, dass das Immunsystem vieler Menschen mit teils heftigen Reaktionen auf die Allergene der Pollenkörner reagiert. Da sich die Anzahl an von Allergien betroffenen Menschen in den letzten Jahrzehnten stetig erhöht hat,^[8] wird sie sogar als „Seuche der modernen Welt“^[9] bezeichnet, welche sich auch zu einer zunehmenden wirtschaftlichen Belastung für Industrienationen entwickelt hat.^[10]

Für betroffene Allergiker ist es für ihr aktuelles Wohlbefinden entscheidend zu wissen, welchen Pollenkörnern sie ausgesetzt sein könnten. Deshalb werden vom Deutschen Wetterdienst neben Wettervorhersagen auch Pollenflugwarnungen veröffentlicht.^[11] Diese basieren darauf, dass Pollenkörner durch eine Pollenfalle kontinuierlich gesammelt und anschließend am folgenden Tag manuell am Mikroskop identifiziert und gezählt werden. Da dieses Verfahren keine aktuellen Daten liefert und einen hohen Personal- und Zeitaufwand erfordert, ist es sinnvoll, nach Alternativen für die Identifizierung von Pollen zu suchen. Eine Möglichkeit, um diese schnell und automatisch erreichen zu können, ist die Anwendung einer on-line Detektion durch spektroskopische oder spektrometrische Methoden in Kombination mit multivariater Statistik.

Die chemische Zusammensetzung, Morphologie und Physiologie von Pollenkörnern wurde bereits in einigen Studien mit Raman- und FTIR-Spektroskopie (Fourier-Transform-Infrarot-Spektroskopie) untersucht.^{[12]-[18]} Letztere beinhaltet auch Anwendungen der oberflächenverstärkten Raman-Streuung (*surface-enhanced Raman scattering*, SERS).^{[19]-[21]} Raman- und FTIR-Spektren dienen dabei nicht nur der Klassifizierung und Identifizierung,^{[12],[13],[16]} sondern auch dazu physiologische Prozesse, wie zum Beispiel die Pollenkeimung^[18] oder den Einfluss von Umwelteinflüssen auf die Pollenchemie^[17] zu verstehen. Bei der Untersuchung von Pollen mit Raman-Spektroskopie tritt häufig ein Fluoreszenzuntergrund in den Spektren auf, der die Interpretation und Auswertung der Daten erschwert.^{[22]-[24]} In SERS-Experimenten werden die Raman-Signale um mehrere Größenordnungen verstärkt^[25] und darüber hinaus auch die Fluoreszenz unterdrückt.

FTIR-Spektren gestatten im Fall einer Untersuchung von Kaliumbromidpresslingen^[16] oder bei Messungen in abgeschwächter Totalreflexion^[26] eine erfolgreiche taxonomische Klassifizierung von Pollen. Für die Klassifizierung von Pollenproben ist allerdings eine Untersuchung einzelner Pollenkörner essentiell. Dies scheiterte bisher an der Überlagerung der Absorptionsspektren durch Streuartefakte, welche die eindeutigen taxonomischen Fingerabdrücke maskieren.^[27]

Eine weitere mögliche Methode für die Klassifizierung von Pollenkörnern, die bereits für andere biologische Proben angewendet wurde, ist die *matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS). Die erhaltenen charakteristischen, fingerabdruckartigen Massenspektren können ebenfalls zur Klassifizierung mit multivariaten Methoden kombiniert werden können.^[28]

Ziel dieser Arbeit ist es, die Informationen aus FTIR-Mikrospektren und SERS- und MALDI-TOF MS-Spektren mit multivariater Auswertung zu kombinieren, um neue Methoden der Klassifizierung von Pollenproben zu erschließen. Dabei soll erprobt werden, ob eine Klassifizierung mit SERS bzw. mit MALDI-TOF MS prinzipiell möglich ist, und ob dabei auch Pollenproben taxonomisch eng verwandter Pflanzenarten differenziert werden können. Darüber hinaus werden FTIR-Spektren einzelner Pollenkörner aufgenommen und dabei untersucht, ob sie durch geeignete mathematische Vorbehandlung trotz der Überlagerung von Absorptions- und Streueffekten eine erfolgreiche taxonomische Klassifizierung gestatten.

Eine zentrale Rolle bei der Datenauswertung nimmt die Erprobung eines Multiblock-Verfahrens ein. Hierbei sollen Daten aus verschiedenen Methoden kombiniert und somit unterschiedliche Bestandteile der komplexen Pollenchemie zusammengeführt werden, um eine verbesserte taxonomische Klassifizierung zu erreichen.

Die hier angewendeten Klassifizierungsmethoden basieren auf Unterschieden der chemischen Zusammensetzung der Pollenproben und könnten deshalb auch zur Untersuchung des Einflusses von Umweltbedingungen auf die Physiologie von Pollen angewendet werden. Deshalb soll im Rahmen dieser Untersuchungen überprüft werden, ob Pollenkörner, die von *cherry leaf roll virus* (CLRV) infizierten Birken stammen, von nicht infizierten Kontrollen unterschieden werden können. Eine Identifizierung von virusinfizierten Pollenproben ist besonders interessant, da bereits gezeigt wurde, dass Viren Pflanzenproteine verändern und damit auch das allergene Potential von Pollenkörnern beeinflussen können.^[29]

Die Beschreibung der Ergebnisse dieser Arbeit wird sich zunächst auf das Potential von SERS zur Charakterisierung und Klassifizierung von Pollenproben konzentrieren (Kapitel 4). Im Vordergrund steht dabei, die SERS-Daten trotz ihrer Komplexität für die taxonomische Klassifizierung zu erschließen. In Kapitel 5 wird gezeigt, dass sich auch MALDI-TOF MS für die Klassifizierung eignet, und es werden multivariate Verfahren zur Unterscheidung taxonomischer Gruppen erprobt. Die Ausnutzung der Kombination von MALDI-TOF MS- und SERS-Daten unter der Verwendung einer *Consensus Principal Component Analysis* (CPCA) wird in Kapitel 6 diskutiert. Kapitel 7 widmet sich der Diskussion von FTIR-Spektren einzelner Pollenkörner, die sich aus der Überlagerung von Absorptions- und

Streuungseffekten zusammensetzen. Im letzten Kapitel 8 werden erstmalig spektroskopische und spektrometrische Daten von virusbefallenen Pollenproben für eine Klassifizierung verwendet.

2. Überblick über die verwendeten Verfahren zur Klassifizierung von Pollenproben

Das Ziel dieser Arbeit besteht darin, Pollenproben mit unterschiedlichen analytischen Methoden zu untersuchen und diese anschließend durch Kombination der erhaltenen Daten mit multivariaten Verfahren zu klassifizieren. Dabei sollen sowohl bekannte und nachweislich für die Untersuchung von biologischen Proben gut geeignete spektroskopisch, mikroskopische und spektrometrische Methoden angewendet, aber auch grundsätzlich neue Methoden erschlossen werden. Ein kurzer Überblick über die in dieser Arbeit verwendeten analytischen (Abschnitt 2.1) und multivariaten Verfahren (Abschnitt 2.2) soll im Folgenden gegeben werden.

2.1 Analytische Verfahren zur Untersuchung von Pollenproben

Methoden der Schwingungsspektroskopie bieten diverse Vorteile für die Klassifizierung und Charakterisierung von biologischen Proben. So können mit geringem präparativem und zeitlichem Aufwand sowohl quantitative als auch qualitative, spezifische Informationen von minimalen Probenmengen gewonnen werden. Da es sich um nicht invasive und nicht destruktive Analysemethoden handelt, können sogar in vivo Untersuchungen^[30] durchgeführt werden. Die hohe Auflösung der Spektren und die im Vergleich zu anderen Methoden wie z.B. der Fluoreszenzspektroskopie relativ schmalen Banden führt zu einem sehr hohen Informationsgehalt, was auch die Untersuchung von komplexen Proben ermöglicht. Handelt es sich um hoch komplexe, biologische Proben können durch die Schwingungsspektroskopie Spektren erhalten werden, bei denen sich die Signale diverser Moleküle überlagern. Diese können als Fingerabdrücke mit den Spektren anderer Proben verglichen werden, um eine Klassifizierung der Proben anhand von spektralen Unterschieden, die ggf. auf komplizierte biologische Zusammenhänge zurückgeführt werden können, zu erreichen. Um charakteristische Spektren zu erhalten, wird sowohl die Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie), als auch die Raman-Spektroskopie angewendet. Diese basieren auf zwei grundsätzlich unterschiedlichen Prozessen, die in Abbildung 2.1 schematisch dargestellt sind: der Absorption von Licht im Infrarotbereich und der inelastischen Streuung von Licht. Allgemein ist bei der Streuung von Licht an einem Molekül der Prozess der elastischen Rayleigh-Streuung am wahrscheinlichsten. Bei der inelastischen Raman-Streuung unterscheidet man zwischen zwei Prozessen: Trifft monochromatisches

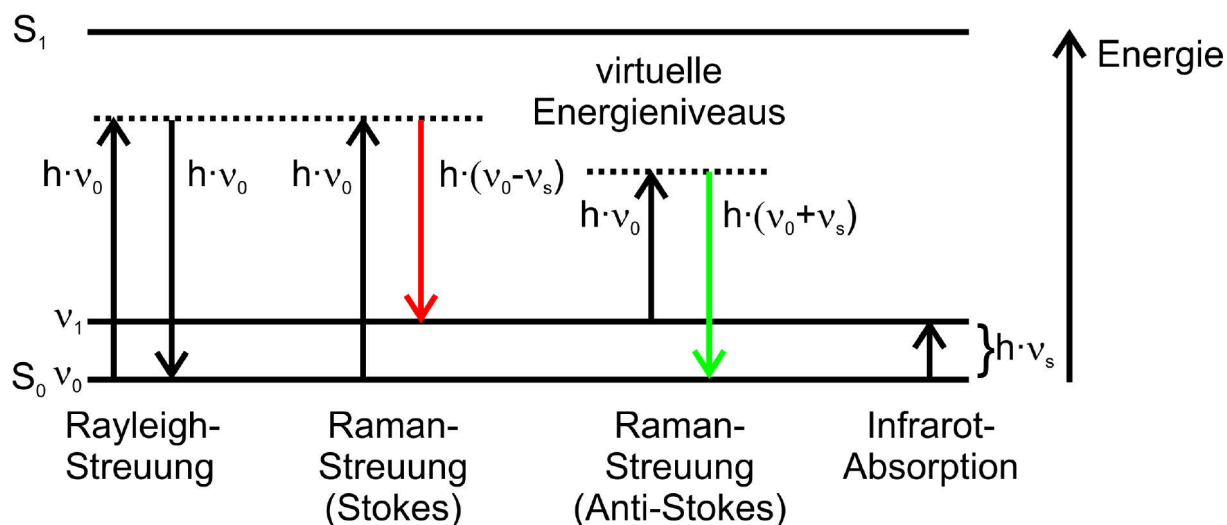


Abbildung 2.1: Schematische Darstellung möglicher Prozesse bei der Wechselwirkung von einem Molekül mit Licht.

Licht auf ein Molekül im Schwingungsgrundzustand v_0 kann die Energie $E_s = h \cdot v_s$ auf das Molekül übertragen werden und man spricht von Stokes-Streuung. Befindet sich das Molekül beim Zusammenstoß in einem angeregten Schwingungszustand, wie z.B. v_1 , kann die Energie $E_s = h \cdot v_s$ auf das Photon übertragen werden und es handelt sich um Anti-Stokes-Streuung. Da für das Auftreten von Anti-Stokes-Raman-Streuung das Molekül im angeregten Schwingungszustand vorliegen muss, was bei Raumtemperatur sehr unwahrscheinlich ist, wird in der Raman-Spektroskopie zumeist die Stokes-Raman-Streuung untersucht.^[31] Aus der Differenz der Energien der inelastisch gestreuten Lichtquanten $E_R = h \cdot (v_0 \pm v_s)$ und der Laserstrahlung $E_0 = h \cdot v_0$, kann die Energie der Schwingungszustände $E = h \cdot v_s$ und damit die Struktur des Moleküls abgeleitet werden. Wird ein Schwingungsspektrum nicht durch Raman-Streuung, sondern durch Absorption von Infrarotlicht erhalten, entspricht die Energie der Schwingungszustände der Energie der absorbierten Strahlung. Da für FTIR- und Raman-Spektroskopie unterschiedliche Auswahlregeln gelten, handelt es sich bei diesen um komplementäre Methoden, die unterschiedliche Informationen über Molekülschwingungen liefern. Prinzipiell sind Schwingungen, bei denen das Dipolmoment des Moleküls verändert wird infrarot-aktiv, ändert sich hingegen die Polarisierbarkeit, so spricht man von Raman-aktiven Schwingungen.

In dieser Arbeit werden die FTIR-Spektroskopie, die Raman-Streuung, sowie die oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS) zur Untersuchung der Molekülschwingungen von Pollenproben angewendet. Die Grundlagen dieser Methoden, sowie der jeweilige historische Kontext deren Entwicklung sollen in den folgenden

Abschnitten kurz erläutert werden. Darüber hinaus werden die Pollenproben auch massenspektrometrisch und mikroskopisch untersucht, weshalb auch ein Überblick über die Methoden der *matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS) und der *environmental scanning electron microscopy* (ESEM) gegeben werden soll.

2.1.1 Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie)

Für die Entwicklung der FTIR-Spektroskopie sind mehrere Ereignisse von grundlegender Bedeutung.^[32] Zunächst ist dabei natürlich die Entdeckung der Infrarotstrahlung durch Sir William Herschel im Jahre 1800 zu nennen. Umfassende Spektralanalysen mit infrarotem Licht konnten ab 1937 durchgeführt werden, da in diesem Jahr das erste vollautomatische Spektralphotometer von Lehrer an der BASF entwickelt wurde. Die entscheidende Weiterentwicklung der Spektrometer gelang jedoch erst Mitte des 20. Jahrhunderts, als das von Michelson im Jahre 1890 entwickelte Interferometer mit der Fourier-Transformation am Computer verknüpft werden konnte und so die ersten kommerziellen FT-Spektrometer erschienen. Diese bieten den entscheidenden Vorteil, dass nicht wie bei dispersiven Spektrometern das Spektrum durch schrittweise Variation der Energie des absorbierten Lichtes erhalten wird, sondern durch Fourier Transformation eines Interferogramms. Dieses wird generiert, indem polychromatisches Licht im Michelson-Interferometer durch einen halbdurchlässigen Spiegel geteilt und anschließend auf einen beweglichen und einen festen Spiegel geleitet wird. Durch Variation der Weglänge des Lichts, das auf den beweglichen Spiegel trifft, ergibt sich ein Phasenunterschied der beiden elektromagnetischen Wellen, der beim erneuten Zusammenführen zur Interferenz führt.

Aufgrund der vielen Vorteile, die die FTIR-Spektroskopie bietet, ist sie momentan als die modernste Methode für die Untersuchung komplexer, biologischer Proben etabliert^[33] und wird neben zahlreichen Anwendungen in der Medizin^[34] z.B. auch für die Klassifizierung von Mikroorganismen^[35] angewendet. Folglich wurden auch Pollenkörner bereits durch diese Methode untersucht. Dabei wurde u.a. die biochemische Zusammensetzung charakterisiert,^[36] der Einfluss von Umweltparametern analysiert^[17] und Pollenproben verschiedener Gattungen taxonomisch klassifiziert^[16].

Im Rahmen dieser Arbeit wurde die FTIR-Spektroskopie mit Licht im mittleren Infrarot (MIR, 4000-400 cm⁻¹) im Transmissionsmodus angewendet, um spezifische, spektrale Fingerabdrücke von Molekülen, die sich in Pollenkörnern befinden, zu erhalten. Zur Untersuchung von individuellen Pollenkörnern ist dabei eine größere Strahlungsintensität als die von

kommerziell erhältlichen Spektrometern notwendig. Zu diesem Zweck kann die FTIR-Mikrospektroskopie an einem Elektronenspeicherring mit der sehr viel intensiveren Synchrotronstrahlung durchgeführt werden.^{[37],[38]} Diese wird dann emittiert, wenn sich ein geladenes Teilchen mit relativistischer Geschwindigkeit in einer Kreisbahn bewegt.

Bei der FTIR-Spektroskopie von Feststoffen kommt es häufig zur Überlagerung des Absorptionsspektrums, aus dem chemische Informationen der Probe geschlossen werden können, mit Streuartefakten, die u.a. brechungsindex- und wellenlängenabhängig sind. Um den Einfluss dieser physikalisch begründeten Artefakte zu verringern, wurde der sogenannte EMSC (*Extended Multiplicative Signal Correction*) Algorithmus entwickelt. Bei diesem wird davon ausgegangen, dass sich bei einem Spektrum z_i die chemischen Informationen von den wellenlängenabhängigen Streuartefakten trennen lassen, indem dieses auf ein ideales Spektrum korrigiert wird. Da ein ideales Spektrum meistens nicht vorhanden ist, wird stattdessen das Mittelwertspektrum m aller im Datensatz vorhandenen Spektren verwendet. Für die EMSC-Korrektur der Spektren wird jeweils nach Gleichung 2.1 ein konstanter Parameter a_i , ein multiplikativer Parameter b_i , sowie ein linearer, wellenlängenabhängiger Parameter d_i und ein quadratischer, ebenfalls wellenlängenabhängiger Parameter e_i bestimmt. Da sich die zu isolierenden, chemischen Informationen in den Residuen ε_i befinden, kann nach Gleichung 2.2 anschließend anhand der erhaltenen Parameter das korrigierte Spektrum $z_{i,Korr}$ bestimmt werden.^{[39],[40]}

$$z_i = a_i + b_i m + d_i \bar{\nu} + e_i \bar{\nu}^2 + \varepsilon_i \quad (2.1)$$

$$z_{i,Korr} = \frac{z_i - a_i - d_i \bar{\nu} - e_i \bar{\nu}^2}{b_i} \quad (2.2)$$

Werden biologische Partikel mit einer Größe von 10 bis 100 μm mit FTIR-Mikrospektroskopie untersucht, so sind die Streuartefakte, die in den Spektren auftreten zu komplex um reine Absorptionsspektren durch eine EMSC-Korrektur zu erhalten. Dies ist darauf zurückzuführen, dass sogenannte Dispersionsartefakte auftreten, die aus resonanter Mie-Streuung an den Partikeln resultieren.^[41] Basierend auf dem EMSC-Algorithmus wurden neue Ansätze erprobt, um auch den Einfluss dieser komplexeren Artefakte auf die Spektren zu reduzieren.^{[42]-[45]} Eine Extraktion der reinen Absorptionsbeiträge aus den FTIR-Spektren von Pollenkörnern ist allerdings momentan noch nicht möglich.^[27] Das Ziel dieser Arbeit ist es, Daten von Pollenkörnern zu erhalten, die trotz der Überlagerung mit Streuartefakten eine erfolgreiche Klassifizierung von Pollenkörnern ermöglichen.

2.1.2 Raman-Streuung

Wird Materie mit Licht bestrahlt, kann es zu verschiedenen Wechselwirkungen kommen. Bei der Lichtstreuung wird die elektromagnetische Welle von ihrer ursprünglichen Ausbreitungsrichtung abgelenkt. Dabei werden die meisten Lichtquanten elastisch, d.h. ohne Energieübertragung an den Molekülen gestreut. In den 20er Jahren des 19ten Jahrhunderts wurde festgestellt, dass auch ein geringer Teil der Lichtquanten inelastisch, d.h. mit Energieübertragung zwischen Molekül und Licht gestreut wird. Dieser Effekt wurde 1923 von Smekal^[46] theoretisch vorhergesagt und 1928 sowohl von Raman und Krishnan, anhand der Streuung von Licht an Flüssigkeiten,^[47] als auch von Landsberg und Mandelstam, anhand der Streuung von Licht an Kristallen,^[48] bewiesen. Für die experimentelle Bestätigung dieses Effektes wurde dieser nach Raman benannt und Raman und Krishnan wurden 1930 mit dem Nobelpreis für Physik ausgezeichnet.

Verglichen mit der FTIR-Spektroskopie bietet die Raman-Spektroskopie sowohl den Vorteil, dass eine höhere räumliche Auflösung erreicht werden kann, als auch, dass wässrige, biologische Proben problemlos untersucht werden können. Dies ist darauf zurückzuführen, dass Wasser intensitätsschwache Banden im Raman-Spektrum aufweist und somit Banden anderer Moleküle nicht überlagert werden. Die Raman-Spektroskopie wurde deshalb bereits in zahlreichen Anwendungen auf die Untersuchung biologischer Proben angewendet.^[49] Auch Pollenproben wurden durch diese schwingungsspektroskopische Methode bereits untersucht. Dabei wurde unter anderem eine taxonomische Klassifizierung,^[13] die Charakterisierung von Carotinoiden,^[15] sowie die Analyse molekularer Veränderungen bei der Pollenkeimung^[18] erreicht.

Die Intensität der Raman-Streuung I_R der j -ten Normalschwingung Q_j ist abhängig von der Polarisierbarkeit α , der Frequenz der Raman-Schwingung $\nu_0 \pm \nu_s$ und einer Konstante μ (Gleichung 2.3).^[50] Um Raman-Intensitäten verschiedener Experimente vergleichen zu können, ist es in analytischen Anwendungen üblich, den Raman-Streuquerschnitt σ_j der jeweiligen Schwingungen, bzw. der jeweiligen Moleküle, empirisch zu bestimmen. Dieser gibt die Wahrscheinlichkeit für das Ereignis an, dass ein Photon mit einer bestimmten Raman-Verschiebung gestreut wird und fasst somit molekül- und experimentspezifische Parameter zusammen. Aus dem Produkt aus Streuquerschnitt, Anregungsintensität I_0 und der Anzahl der Moleküle N ist somit die Raman-Intensität einfach abschätzbar (Gleichung 2.3).

$$I_R = \mu(\nu_0 \pm \nu_s)^4 \alpha_j^2 Q_j^2 \approx I_0 \sigma_j N \quad (2.3)$$

Im Vergleich zu anderen physikalischen Prozessen, wie z.B. der Fluoreszenz, ist die Raman-Streuung recht unwahrscheinlich, weshalb sowohl die Werte für den Streuquerschnitt σ_j (ca. 10^{-30} cm^{-1}), als auch die der Raman-Intensitäten vergleichsweise niedrig sind. Deshalb müssen Raman-Experimente mit einer hohen Anregungsintensität, einer langen Messzeit oder einer hohen Anzahl an Molekülen durchgeführt werden. Raman-Signale können allerdings auch dann nur detektiert werden, wenn keine intensiven Fluoreszenzsignale auftreten und diese überlagern. Eine Möglichkeit fluoreszenzfreie, intensive Raman-Signale zu detektieren, bietet die oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS).

2.1.3 Oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS)

Das Phänomen der oberflächenverstärkten Raman-Streuung wurde erstmalig 1974 von Fleischmann *et al.*^[51] bei der Untersuchung von Pyridin an einer aufgerauten Silberelektrode beobachtet. Die signifikant höheren Raman-Signale wurden allerdings lediglich auf Adsorptionseffekte zurückgeführt. Erst drei Jahre später wurde festgestellt, dass eine Signal-Verstärkung dieses Ausmaßes nur durch einen weiteren Effekt erklärt werden kann^[52] und es wurde erstmalig von Oberflächenverstärkung durch Feldverstärkung gesprochen^[53].

Heute versteht man unter SERS den Effekt, dass Moleküle in unmittelbarer Umgebung von Metallnanostrukturen Raman-Streuquerschnitte von bis zu 10^{-16} cm^{-1} aufweisen^[25] und damit Raman-Signale detektiert werden können, die um bis zu acht Größenordnungen intensiver sind als die nicht verstärkten Signale. Man führt dies auf zwei verschiedene Mechanismen zurück: die chemische und die elektromagnetische Verstärkung.

Der wesentlich kleinere Anteil der Verstärkung wird auf die chemische Verstärkung zurückgeführt. Sie wird durch die Wechselwirkung des Moleküls mit der Metalloberfläche erklärt, wodurch der Raman-Streuquerschnitt verändert wird.^[54] Aufgrund dieser Wechselwirkung wird eine Ladungsverschiebung hervorgerufen, die die Polarisierbarkeit des Moleküls und damit das induzierte Dipolmoment beeinflusst. Außerdem können auch Charge-Transfer Zustände entstehen,^[55] die ebenfalls die Wechselwirkung des Moleküls mit Licht beeinflussen. Durch die Veränderungen der elektronischen Struktur des untersuchten Moleküls kann das SERS-Spektrum andere Intensitätsverhältnisse oder sogar unterschiedliche Banden oder Bandenpositionen als das normale Raman-Spektrum aufweisen. Außerdem wird dadurch die Lebensdauer von elektronisch angeregten Zuständen verkürzt, so dass in SERS-Spektren häufig ein geringerer Fluoreszenzuntergrund als in normalen Raman-Spektren auftritt.^[56]

Die elektromagnetische Verstärkung basiert auf der Wechselwirkung des eingestrahnten und des gestreuten Lichts mit Oberflächenplasmonen. Wird eine Metallnanostruktur mit Licht der Energie $E_0(\nu_0)=h\cdot\nu_0$ bestrahlt, resultiert daraus eine kollektive Schwingung der Leitungselektronen (Plasmonen), wenn diese in Resonanz zu ν_0 sind. Dies führt dazu, dass ein Dipol und damit ein elektrisches Feld $E_{ind}(\nu_0)$ induziert wird. Befindet sich ein Molekül in der unmittelbaren Umgebung der Metallnanostruktur, so wirkt die Summe der beiden Felder $E_{ind}(\nu_0) + E_0(\nu_0)$ auf dieses. Befindet sich das elektromagnetische Feld der Raman-Streustrahlung $E_R(\nu_R)$ ebenfalls in Resonanz mit den Oberflächenplasmonen, so wird auch durch dieses ein Dipol und damit ein elektromagnetisches Feld induziert. Es wird somit sowohl das elektromagnetische Feld der Anregungsstrahlung, als auch das der Raman-Streustrahlung durch Metallnanostrukturen verstärkt.

Die Intensität des SERS-Signals ist somit nicht nur abhängig von der Anregungsintensität I_0 und der Anzahl der Moleküle N , sondern auch von der chemischen Verstärkung, die den Streuquerschnitt der adsorbierten Probe (σ_{SERS}) beeinflusst, und der elektromagnetischen Verstärkung, die sich aus der jeweils quadratischen Verstärkung des Anregungsfeldes ($A(\nu_0)$) und des Raman-Streufeldes ($A(\nu_R)$) zusammensetzt (Gleichung 2.4).

$$I_{SERS}(\nu_0) = I_0 N A(\nu_0)^2 A(\nu_R)^2 \sigma_{SERS} \quad (2.4)$$

Die elektrische Feldstärke E und damit auch die Verstärkung A der elektrischen Felder hängt vom Abstand r zwischen Nanostruktur und Analyten in der dritten Potenz ab. Unter der Annahme einer ähnlichen Verstärkung von Anregungs- und Streufeld, lässt sich aus Gleichung 2.4 somit schließen, dass die SERS-Verstärkung mit der zwölften Potenz dieses Abstandes abnimmt (Gleichung 2.5).

$$I_{SERS} \sim \frac{1}{r^{12}} \quad (2.5)$$

Deshalb treten in SERS-Spektren häufig hohe Fluktuationen der Signalintensitäten auf, was durch das Auftreten sogenannter „Hotspots“, also Bereichen sehr hoher Verstärkung zwischen Metallnanostrukturen, noch verstärkt wird.

Aufgrund der Vorteile, die SERS für die Untersuchung biologischer Proben bietet, wurden bereits zahlreiche, biologische Fragestellungen durch diese Methode untersucht.^[57] Die geringe Reproduzierbarkeit der SERS-Daten ist jedoch häufig eine Herausforderung, wenn diese mit multivariaten Methoden für eine Klassifizierung kombiniert werden. Trotzdem

wurden beispielsweise Viren^[58] oder Bakterien^[59] bereits erfolgreich anhand ihrer SERS-Spektren klassifiziert.

2.1.4 *Matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS)

Die Massenspektrometrie ist einer der am häufigsten angewendeten Verfahren zur Untersuchung von analytischen Fragestellungen. Der Analyt wird dabei in die Gasphase überführt und ionisiert, seine Bestandteile nach dem Masse-zu-Ladungs-Verhältnis m/z aufgetrennt und diese anschließend qualitativ und quantitativ erfasst. Bei der Untersuchung von komplexen, biologischen Proben kann das Massenspektrum eine Vielzahl an Peaks unterschiedlicher Bestandteilen enthalten, die, ähnlich wie bei der Schwingungsspektroskopie, als Fingerabdruck der chemischen Zusammensetzung der biologischen Probe dienen können.

Die Grundlagen der Massenspektrometrie wurden Anfang des 20. Jahrhunderts von Sir J. J. Thompson gelegt, als dieser erstmalig das Masse-zu-Ladungs-Verhältnis von Ionen bestimmte. Diese wurden in Vakuumröhren generiert und anschließend durch magnetische und elektrische Felder getrennt und auf einem Fluoreszenzschirm detektiert. Ein Schüler von Thompson, Francis W. Aston, entwarf später das erste Massenspektrometer, das Ionen nach ihrer Masse trennte und anhand ihrer Geschwindigkeit fokussierte. Thompson wurde 1908 mit dem Nobelpreis für Physik und Aston 1922 mit dem für Chemie ausgezeichnet.

Ein Massenspektrometer besteht prinzipiell aus drei Teilen: In der Ionenquelle werden Ionen erzeugt, die dann im Analysator getrennt und am Detektor erfasst werden. In dieser Arbeit wurde als Ionenquelle die Matrix-unterstützte Laserdesorption/Ionisation (*matrix-assisted laser desorption/ionisation*, MALDI), als Analysator ein Flugzeitmassenspektrometer (*time-of-flight mass spectrometer*, TOF MS) und Mikrokanalplatten (*microchannel plate*, MCP) als Detektor verwendet. Die Funktionsweise dieser drei Bestandteile soll im Folgenden kurz erläutert werden.^[60]

Das Grundprinzip von MALDI beruht auf der Kokristallisation des zu untersuchenden Analyten mit einer stark lichtabsorbierenden Matrix, die im Überschuss vorliegt. Häufig werden als Matrix kleine organische Moleküle verwendet, die ultraviolette Licht absorbieren. Anschließend werden die Kokristalle mit kurzen Pulsen (3-10 ns) eines Lasers, beispielsweise eines frequenzverdreifachten Nd:Yag-Lasers ($\lambda=355\text{nm}$), bestrahlt, was zu einer Energieaufnahme durch die Matrix und damit auch zur Verdampfung und Ionisation des Analyten führt.

Die auf diese Weise gebildeten Ionen werden anschließend in einem TOF MS anhand des Masse-zu-Ladungs-Verhältnisses m/z getrennt. Entscheidend für die Trennung der verschiedenen Ionen ist dabei die Flugzeit t , die dieses Ion für die festgelegte Strecke s nach Anlegen einer konstanten Beschleunigungsspannung U in feldfreier Umgebung benötigt. Unter Berücksichtigung der Elementarladung e ergibt sich für die Flugzeit t eines Ions mit Masse m_i :^[60]

$$t = \frac{s}{\sqrt{2eU}} \sqrt{\frac{m_i}{z}} \quad (2.7)$$

Die Ionen werden so anhand unterschiedlicher Flugzeiten getrennt. Nach Gleichung 2.7 können anschließend die entsprechenden Masse-zu-Ladungs-Verhältnisse aus den jeweiligen Flugzeiten bestimmt werden. Um die Flugzeiten bestimmen zu können, muss jeweils ein Start- und ein Endpuls für die Zeiterfassung festgestellt werden. Der Startpuls ergibt sich bei MALDI-TOF MS durch die kurzen Laserpulse, welche die Ionisierung bewirken, während der Endpuls durch Strommessungen am Detektor erfolgt. Dazu müssen ankommende Ionen in ein ausreichend intensives Signal umgewandelt werden. Dies kann zum Beispiel durch Mikrokanalplatten (MCP) erfolgen, welche auf dem Prinzip des Sekundärelektronenvervielfachens basieren. Dabei werden 10^4 bis 10^7 Kanäle aus Halbleitermaterial hintereinander angeordnet. Trifft ein Ion auf einen solchen Kanal, so werden mehrere Elektronen ausgelöst, die durch Anlegen einer Spannung weiter beschleunigt werden. Die Kollision dieser Elektronen mit nachfolgenden Kanälen führt dazu, dass immer mehr Elektronen frei werden, die anschließend durch Messung der Stromstärke detektiert werden können.

Der wesentliche Vorteil von MALDI-TOF MS gegenüber anderen massenspektrometrischen Methoden liegt darin, dass auch sehr große Moleküle, wie zum Beispiel Biopolymere, weitestgehend fragmentierungsfrei analysiert werden können. Deshalb stellt MALDI-TOF MS heutzutage ein wichtiges analytisches Werkzeug für die massenspektrometrische Untersuchung von komplexen biologischen Proben dar, die von menschlichen Gewebeproben^[61] über Pilze^[62] und Bakterien^[63] bis hin zu ganzen Organismen^[64] reichen.

2.1.5 Environmental scanning electron microscopy (ESEM)

Da bei der Untersuchung von Pollenproben durch FTIR-Spektroskopie Daten erhalten werden, die aus einer Überlagerung von Absorptions- und Streubeiträgen resultieren, ist es sinnvoll Größe, Form und Morphologie der Pollenkörner anhand von mikroskopischen Aufnahmen in die Untersuchung der Proben mit einzubeziehen. Dies wurde durch ESEM-

Aufnahmen der Pollenkörner erreicht. Bei ESEM handelt es sich um eine spezielle Variante der Rasterelektronenmikroskopie (*Scanning Electron Microscopy*, SEM), die in den achtziger Jahren des letzten Jahrhunderts von Danilatos entwickelt wurde.^[65] Diese eignet sich besonders für die Untersuchung von biologischen Proben,^[66] da hier nicht im Hochvakuum, sondern nur bei verringertem Druck gearbeitet werden kann. Bei der Elektronenmikroskopie wird allgemein ein fokussierter Elektronenstrahl dazu verwendet, um die Probe zu scannen und dabei verschiedene detektierbare Prozesse, wie zum Beispiel das Herausschlagen von Sekundärelektronen, auszulösen. Diese treffen dann auf die in der Probenkammer befindlichen Gasmoleküle und ionisieren diese. Dabei werden Elektronen gebildet, die zu einer Signalverstärkung führen und positiv geladene Gasmoleküle, welche verhindern, dass die Probe aufgeladen wird, so dass eine Bedampfung mit Kohlenstoff oder Gold nicht notwendig ist. Die Kollision des Elektronenstrahls mit Gasmolekülen führt jedoch dazu, dass der Elektronenstrahl defokussiert wird, was in einer geringeren Auflösung von ESEM verglichen mit SEM resultiert. Da es sich bei Pollenkörnern allerdings um relativ große biologische Proben handelt, ist das Auflösungsvermögen von ESEM ausreichend, um Größe, Form und Morphologie von diesen zu untersuchen.

2.2 Multivariate Verfahren

Da die Auswertung der Daten durch multivariate Verfahren ein wesentlicher Teil dieser Arbeit ist, sollen nun im Folgenden die verwendeten Methoden erläutert und hinsichtlich deren Möglichkeiten zur Auswertung analytischer Pollendaten bewertet werden.

2.2.1 Hierarchische Clusteranalyse (HCA)

Bei der hierarchischen Clusteranalyse (HCA) handelt es sich um ein unüberwachtes Verfahren, bei dem Daten anhand von Ähnlichkeiten in Gruppen zusammengefasst werden. Dabei werden die einzelnen Datenblöcke, die in dieser Arbeit den Spektren der verschiedenen analytischen Methoden entsprechen, als Vektoren im n -dimensionalen Raum dargestellt, wobei jede Dimension jeweils einem x -Wert im Spektrum entspricht. Die Ähnlichkeit zweier Spektren $a(a_1, a_2, \dots, a_n)$ und $b(b_1, b_2, \dots, b_n)$ kann auf Basis der quadrierten euklidischen Distanzen d_{ab} bestimmt werden:

$$d_{ab} = \sum_{i=1}^n (a_i - b_i)^2 \quad (2.8)$$

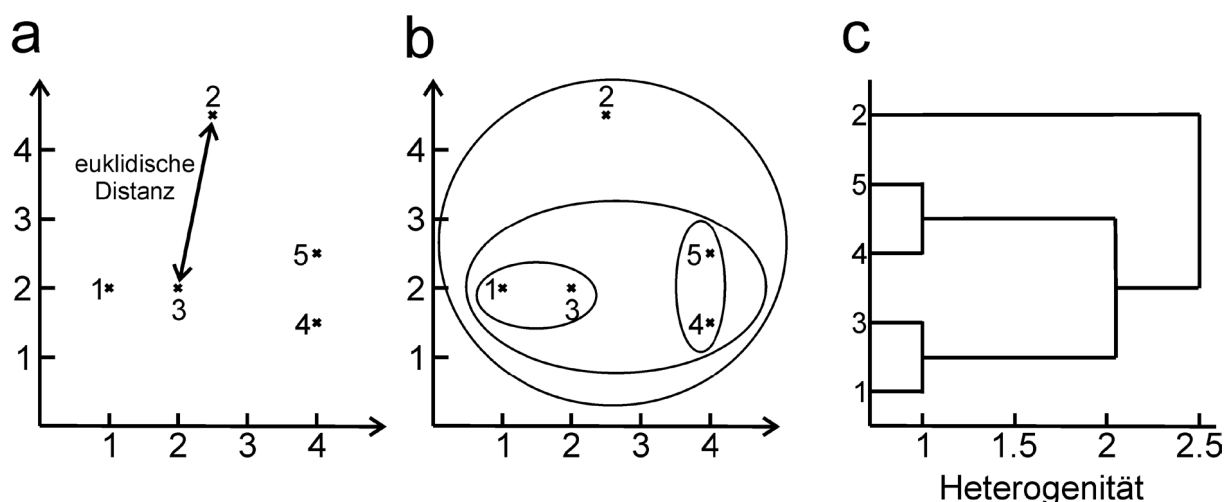


Abbildung 2.2: Clusteranalyse der Punkte 1-5 nach [67].

Anhand der geringsten Distanz werden bei der HCA als Erstes die beiden Objekte mit der geringsten Heterogenität in einen Cluster zusammengefasst. Anschließend werden die Abstände des neu gebildeten Clusters zu den übrigen Vektoren bestimmt und dieser Vorgang solange wiederholt, bis sich alle Spektren in einem Cluster befinden. Der gewählte Algorithmus bestimmt dabei, auf welche Weise die Spektren zu Clustern zusammengefasst werden. Im Rahmen dieser Arbeit wurde der Ward's Algorithmus verwendet. Dieser simuliert jeweils die Aufnahme eines Objektes in einen Cluster und ermittelt dabei den sogenannten Heterogenitätsfaktor, der beschreibt wie homogen, d.h. ähnlich, die Objekte innerhalb dieses neu gebildeten Clusters sind. Das Objekt wird anschließend nur dann in den Cluster integriert, wenn es sich um ausreichend ähnliche Objekte handelt. Das Ergebnis der Clusteranalyse wird anschließend in einem Dendrogramm dargestellt. Durch den Abstand der vertikalen Linien lässt sich daraus die Ähnlichkeit der jeweiligen Spektren und Cluster ablesen. Abbildung 2.2 zeigt exemplarisch, wie die zweidimensionalen Punkte 1-5 anhand der euklidischen Distanzen (Abbildung 2.2a) zu Clustern zusammengefasst werden (Abbildung 2.2b) und anschließend das Ergebnis der Clusteranalyse als Dendrogramm dargestellt wird (Abbildung 2.2c).

Bei der Clusteranalyse handelt es sich also um ein multivariates Verfahren, durch das Ähnlichkeiten und Unterschiede von Spektren anhand spektraler Distanzen analysiert werden können. Bei der Untersuchung komplexer analytischer Pollendaten könnte besonders die Darstellung möglichst homogener Cluster als Dendrogramm dazu führen, dass z.B. taxonomische Verwandtschaftsverhältnisse aus den Ergebnissen abgelesen werden können. Dies wäre dann der Fall, wenn taxonomisch eng verwandte Gruppen auch ähnliche spektrale Distanzen aufweisen, während nur fern verwandte Gruppen durch größere Unterschiede dieser gekennzeichnet sind. Sind diese taxonomischen Ähnlichkeiten und Varianzen jedoch

mit intensiven, taxonomisch irrelevanten Informationen überlagert, so ist eine Untersuchung einfacher spektraler Distanzen nicht ausreichend um eine taxonomische Klassifizierung zu erreichen.

2.2.2 Hauptkomponentenanalyse (*Principal Component Analysis, PCA*)

Die Hauptkomponentenanalyse (*Principal Component Analysis, PCA*) wurde im Jahre 1901 erstmalig von Karl Pearson formuliert^[68] und zunächst vor allem in der Psychologie unter dem Namen „Faktorenanalyse“ verwendet. Mit der Weiterentwicklung von Computern wurde dieses multivariate Verfahren in verschiedensten Disziplinen, u.a. in Sozial- und Naturwissenschaften eingesetzt.^[40]

Die PCA kann angewendet werden, um Unterschiede in komplexen spektralen Daten herauszuarbeiten und diese ohne störende, redundante Informationen zu analysieren. Können diese Unterschiede auf Klassenunterschiede, beispielsweise von taxonomischen Gruppen von Pollen, zurückgeführt werden, so ist die PCA ein leistungsstarkes Hilfsmittel für die Klassifizierung von Pollenspektren.

Durch die PCA wird eine Varianzgewichtung und eine Reduktion der Daten erreicht. Durch Linearkombination der ursprünglichen, unterschiedlich gewichteten Daten werden sogenannte latente Variablen oder Hauptkomponenten erzeugt. Diese bilden ein neues Koordinatensystem, welches darauf ausgerichtet ist, die maximal mögliche Varianz der Originaldaten abzubilden. Deshalb weist die erste Hauptkomponente in Richtung der maximalen Varianz, und die weiteren Hauptkomponenten stehen orthogonal zu dieser und beinhalten jeweils den maximalen Anteil der restlichen Varianz.

Mathematisch wird bei der PCA die ursprüngliche Datenmatrix X in eine Scoresmatrix T und eine Faktorenmatrix P zerlegt (Abbildung 2.3). Die Scores in der Matrix T beinhalten dabei

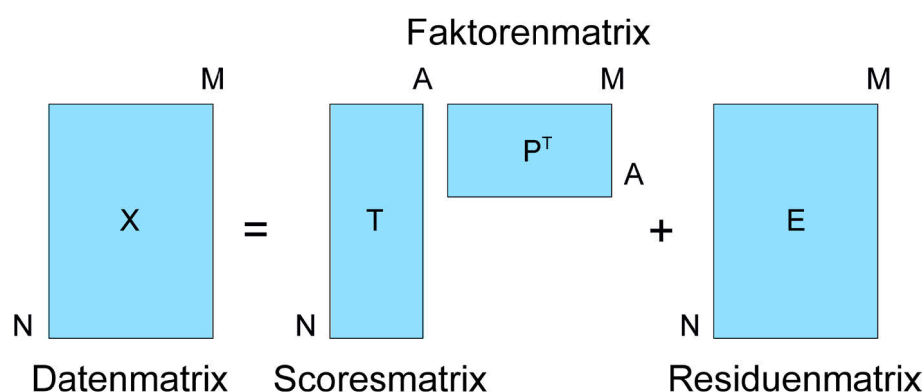


Abbildung 2.3: Matrizen der Hauptkomponentenanalyse nach [40].

die neuen Koordinatenwerte, während die Loadings, die sich in der Faktorenmatrix befinden, den Zusammenhang des ursprünglichen mit dem Hauptkomponentenkoordinatensystem herstellen. Die Residuenmatrix E enthält die Differenz zwischen ursprünglicher Datenmatrix und den über Scores und Loadings reproduzierten Daten und damit den Teil der Originaldaten, der varianzbezüglich nur wenig relevant ist. Besteht die Datenmatrix aus N Spektren mit jeweils M Ordinatenwerten, so werden bei der PCA A Hauptkomponenten gebildet, die jeweils N Scores und M Loadings aufweisen (Abbildung 2.3).

2.2.3 *Partial least squares regression (PLSR)*

Bei der *partial least squares regression* (PLSR) handelt es sich um eine multivariate Regressionsmethode. Erste Grundlagen dafür wurden in den frühen 1970er Jahren von Herman Wold gelegt, als dieser begann einen Algorithmus zu entwickeln, um ökonomische Daten auszuwerten.^[69] Anschließend wurde die PLSR auch in anderen Fachbereichen, u.a. in der Chemie^[70], verwendet und, v.a. von Harald Martens^[71], kontinuierlich weiterentwickelt. Heutzutage wird die PLSR in vielen unterschiedlichen Bereichen angewendet, um komplexe Daten zu analysieren.^[40]

Die PLSR ist ein multivariates Verfahren, bei dem die Zielgröße der Klassifizierung bei der Auswertung mit einbezogen wird. Dies stellt den entscheidenden Unterschied zu den bisher vorgestellten multivariaten Verfahren der PCA und HCA dar. Der Vorteil dieser Methode für die Auswertung von Pollenspektren liegt darin, dass die entscheidenden, ggf. kleinen Unterschiede zwischen Gruppen herausgearbeitet werden. Diese können von den großen Unterschieden in der Untersuchung euklidischer Distanzen oder Unterschieden, die durch die Varianzen im Datensatz gegeben sind, abweichen.

Das Ziel der PLSR besteht darin, den Zusammenhang zwischen Messgrößen X , die in dieser Arbeit den Spektren entsprechen, und Zielgrößen y bei denen es sich z.B. um taxonomische Gruppen handelt, herzustellen. Man unterscheidet dabei zwischen PLSR1 und PLSR2, je nachdem ob eine oder mehrere Zielgrößen, d.h. ein Zielgrößenvektor oder eine Zielgrößenmatrix vorliegen. Da in dieser Arbeit jeweils nur eine Zielgröße verwendet wurde, soll im Folgenden die Funktionsweise der PLSR1 erläutert werden.

Prinzipiell wird bei der PLSR sowohl eine PCA der X -Daten, als auch eine PCA der y -Daten durchgeführt. Diese sind allerdings nicht unabhängig voneinander, sondern jeweils beeinflusst durch die Ergebnisse der anderen PCA. Abbildung 2.4 zeigt schematisch den mathematischen Hintergrund einer PLSR1. Die PCA der X -Daten ergibt die Scoresmatrix T und die spektrale Loadingsmatrix P , während die des y -Vektors die Scoresmatrix U und die chemische

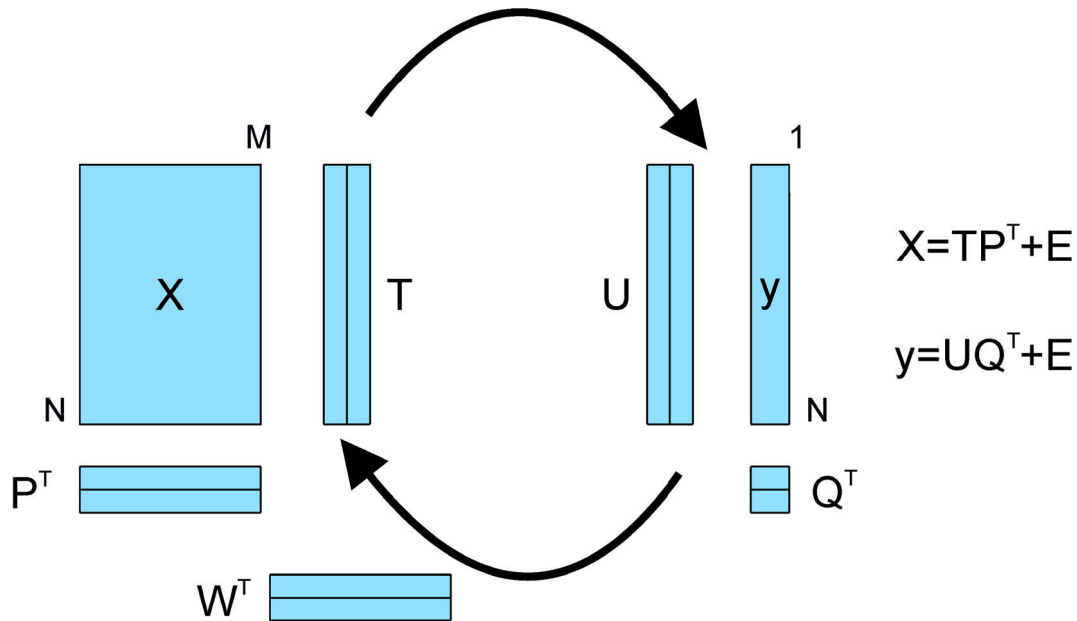


Abbildung 2.4: Schematische Darstellung der mathematischen Grundlage der PLS1.

Loadingsmatrix Q ergibt. Als Zwischenschritt wird eine weitere Matrix W benötigt. Diese wird als gewichtete Loadingsmatrix bezeichnet und beinhaltet die Verbindung zwischen X und y -Daten. Im Folgenden soll exemplarisch anhand der Berechnung der a -ten PLSR-Komponente gezeigt werden, wie aus den X - und y -Daten ein PLSR-Modell erstellt wird. Für die erste PLSR-Komponente wird dabei für X_a die Datenmatrix X und für y_a der Zielgrößenvektor y verwendet. Der y -Vektor wird zunächst als erste Schätzung für die Scores t verwendet und mit der Methode der kleinsten Fehlerquadrate die gewichteten Loadings w_a^T aus folgender Gleichung bestimmt:

$$X_a = y_a w_a^T + E \quad (2.9)$$

Die gewichteten Loadings müssen dabei die Nebenbedingungen erfüllen, dass sie orthogonal zueinander sind und jeweils die Länge 1 aufweisen. Mit Hilfe der gewichteten Loadings können, erneut durch die Methode der kleinsten Fehlerquadrate, die Scores t_a und die spektralen Loadings p_a^T bestimmt werden:

$$X_a = t_a w_a^T + E \quad (2.10)$$

$$X_a = t_a p_a^T + E \quad (2.11)$$

Anschließend werden die chemischen Loadings q_a berechnet, indem der y -Vektor auf die Scores t_a regressiert wird:

$$y_a = t_a q_a + f \quad (2.12)$$

Damit sind alle Parameter der PLSR-Komponente bestimmt und die Informationen, die in dieser enthalten sind, werden aus den X- und y-Daten für die Berechnung der weiteren PLSR-Komponenten subtrahiert:

$$X_{a+1} = X_a - t_a p_a^T \quad (2.13)$$

$$y_{a+1} = y_a - t_a q_a \quad (2.14)$$

Anschließend werden die Gleichungen 2.9 bis 2.14 auf die neu erhaltenen Daten solange erneut angewendet, bis die Anzahl A_{\max} an PLSR-Komponenten berechnet ist. A_{\max} wird dabei durch eine Validierung bestimmt.

Das Ziel der PLSR-Regression ist das Vorhersagen der Zielgröße y durch ein Spektrum X anhand der Regressionskoeffizienten b_0 und b (Gleichung 2.15). Diese lassen sich nach den Gleichungen 2.16 und 2.17 nun nach den obigen Gleichungen durch die Loadings der PLSR-Komponenten ausdrücken, wobei \bar{x}^T und \bar{y} die Mittelwerte aller in die Kalibrierung eingehenden X- und y-Werte darstellen.^[40]

$$y = b_0 + Xb \quad (2.15)$$

$$b = W(P^T W)^{-1} q \quad (2.16)$$

$$b_0 = \bar{y} - \bar{x}^T b \quad (2.17)$$

2.2.4 Consensus Principal Component Analysis (CPCA)

Bei der CPCA handelt es sich um eine Weiterentwicklung der PCA (siehe Abschnitt 2.2.2), die entwickelt wurde, um Datensätze verschiedener Methoden zu kombinieren. Es wurde gezeigt, dass diese Multiblock-Methode dazu fähig ist, FTIR-, NIR- und Raman-spektroskopischen Daten sowohl miteinander^[72] als auch mit nicht spektroskopisch erfassten Probeneigenschaften^[73] zu verknüpfen.

Im Rahmen dieser Arbeit wurden damit MALDI-TOF MS-Daten mit schwingungsspektroskopischen Daten kombiniert, um Pollenproben zu klassifizieren. Da es sich bei diesen um Daten komplementärer Methoden handelt, können auf diese Weise die vorhandenen Informationen verknüpft werden, um einen möglichst großen Teil der komplexen Chemie von Pollenproben abzubilden und bei der Klassifizierung mit einzubeziehen.

Bei der CPCA werden die Daten der einzelnen Methoden X_A^b , die auch als Blöcke bezeichnet werden, in einen sogenannten globalen Block X zusammengefügt. Generell wird bei der CPCA sowohl eine PCA des globalen, als auch der einzelnen Blöcke durchgeführt: Ähnlich wie bei der PLSR (siehe Abschnitt 2.2.3) sind diese unterschiedlichen PCAs jedoch abhängig voneinander. Aus dem Datensatz der globalen Daten X werden A Hauptkomponenten, die aus den globalen Loadings p_a^T und den globalen Scores t_a bestehen, gebildet:

$$X = t_a p_a^T + E_a \quad (2.18)$$

Äquivalent wird jeweils aus den Daten X_A^b des Blockes b A Hauptkomponenten aus Blockloadings p_a^{bT} und Blockscores t_a^b erzeugt:

$$X_A^b = t_a^b p_a^{bT} + E_a^b \quad (2.19)$$

Der Algorithmus der CPCA bildet dabei die Loadings und Scores der verschiedenen Hauptkomponenten und Blöcken in 2 Schritten: der Parameterbestimmung und der Deflation. Bei der Parameterbestimmung werden jeweils die Parameter des globalen Blocks und der einzelnen Blöcke gebildet. Dies geschieht in einem iterativen Prozess der im Folgenden für

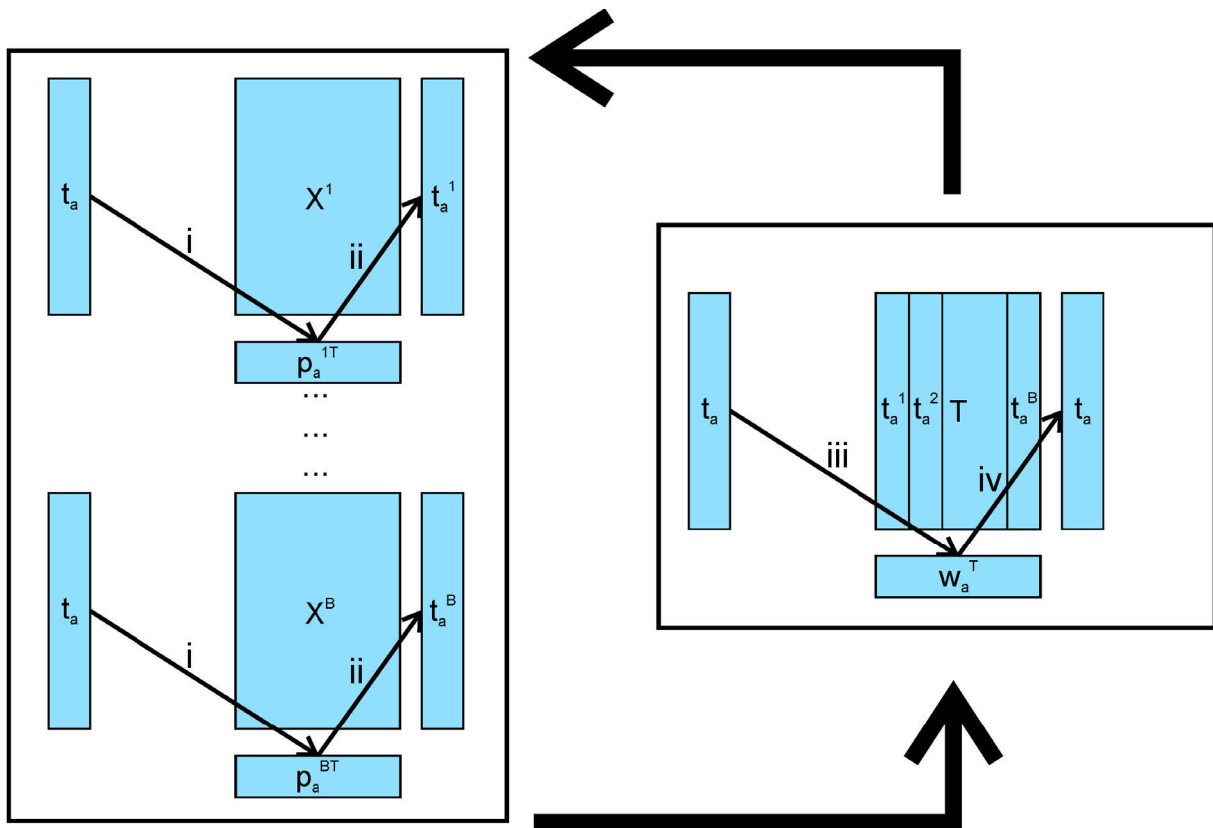


Abbildung 2.5: Schematische Darstellung des iterativen Verfahrens zur Bestimmung der CPCA-Parameter nach [74].

die a -te CPCA-Komponente exemplarisch gezeigt werden soll (Abbildung 2.5). Als Erstes wird ein willkürlicher Vektor t_a als globaler Scoresvektor gewählt. Dieser wird anschließend verwendet, um mit spaltenweiser Projektion auf den jeweiligen Datenblock X^1 bis X^B die Blockloadings p_a^{1T} bis p_a^{BT} zu bestimmen (Abbildung 2.5 i). Aus diesen werden danach zeilenweise die Blockscores t_a^1 bis t_a^B entwickelt (Abbildung 2.5 ii), welche zusammengesetzt die Blockscoresmatrix T ergeben (Abbildung 2.5 rechts). T wird anschließend auf die globalen Scores t_a projiziert, um die globalen, gewichteten Loadings w_a^T zu erhalten (Abbildung 2.5 iii). Diese werden dann verwendet, um eine neue Schätzung der globalen Scores t_a zu bestimmen (Abbildung 2.5 iv), mit denen dann der komplette Prozess bis zur Konvergenz von t_a wiederholt wird. Ist Konvergenz erreicht, also die Parameter der a -ten CPCA-Komponente bestimmt, wird die Varianz, die in der entsprechenden Komponente beschrieben ist vom Datensatz subtrahiert und der so gebildete, deflatierte Datensatz für die Berechnung der weiteren CPCA-Komponenten verwendet.

2.2.5 Künstliche neuronale Netze (*artificial neural networks*, ANN)

Künstliche neuronale Netze (*artificial neural networks*, ANN) sind Mustererkennungsverfahren, die auf einer computerbasierten Nachbildung der Funktionsweise des menschlichen Gehirns aufgebaut sind. Die ersten Netzwerke, die auf Neuronen basierten und zum Lösen von arithmetischen und logischen Funktionen eingesetzt werden konnten, wurden bereits 1943 von Warren McCulloch und Walter Pitts beschrieben.^[75] Vier Jahre später erkannten diese auch das Potenzial dieser Netzwerke für die Erkennung räumlicher Muster.^[76] Zu diesem Zweck wurde 1958 auch der erste Neurocomputer am MIT von Frank Rosenberg und Charles Wightman entwickelt. Diese entwickelten dafür das sogenannte Perzeptron-Modell^[77], das bis heute die Grundlage künstlicher neuronaler Netze bildet. Entscheidend für die Weiterentwicklung dieses Modells war die Entwicklung des Backpropagation Algorithmus[‘],^{[78],[79]} der ein schnelles und robustes Lernen mehrstufiger Netze mit Gradientenabstiegsverfahren des Netzwerkfehlers ermöglicht.^[80] Heutzutage werden ANN in den unterschiedlichsten Fachbereichen für die Mustererkennung und Lösung komplexer Fragestellungen eingesetzt, so auch z.B. für die Klassifizierung von FTIR-^[35] und SERS-Spektren^[81] biologischer Proben. Dabei wird eine große Datenmenge für das Trainieren, Validieren und Testen des Netzes benötigt. Ist diese Menge an Daten vorhanden, so können durch das ANN komplexe, nicht lineare Zusammenhänge aufgedeckt werden, die mit anderen Verfahren oft nicht untersucht werden können. ANN zeichnen sich dabei durch eine besondere Robustheit gegen Störungen, wie zum Beispiel Rauschen oder irrelevante Informationen im Datensatz aus.

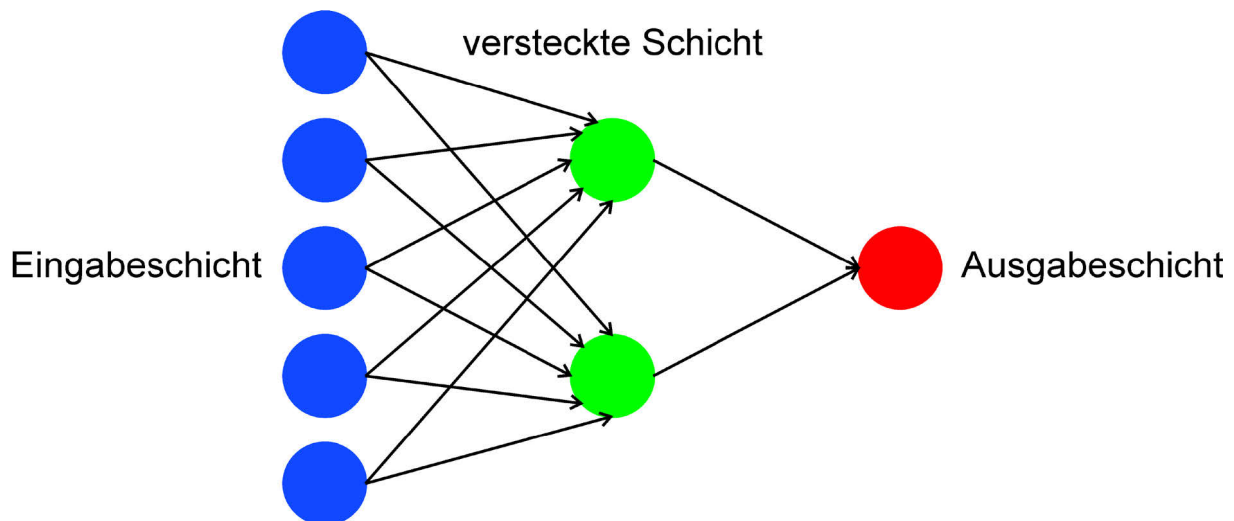


Abbildung 2.6: Beispiel für die Topologie eines vorwärtsgerichteten, zweischichtigen, neuronalen Netzes.

Die Grundlage künstlicher neuronaler Netze bildet das Verknüpfen mehrerer Einheiten (Neuronen) zu einem Netzwerk. Die Struktur dieses Netzes, die auch Topologie genannt wird, wird als Erstes, bevor mit dem Training des Netzes begonnen wird, festgelegt. Im Rahmen dieser Arbeit werden vorwärtsgerichtete, zweischichtige neuronale Netze verwendet. Dies bedeutet, dass jede Schicht ausschließlich mit der jeweils folgenden Schicht verbunden ist, und auch eine verdeckte Schicht des Netzes gebildet wird, deren Zweck lediglich in der Informationsverarbeitung, nicht aber der Betrachtung und Auswertung durch den Nutzer liegt (Abbildung 2.6). Ist die Topologie festgelegt, wird die Gewichtung der einzelnen Einheiten in der Trainingsphase optimiert, um das vorgegebene Ergebnis, das in der Ausgabeschicht für den Nutzer sichtbar ist, möglichst korrekt wiederzugeben. Um dies zu erreichen, wird der Backpropagation Algorithmus angewendet. Dieser bestimmt jeweils für die gewichteten Eingabeschichten einen quadratischen Fehler zwischen vom ANN vorhergesagter und angestrebter, vorgegebener Ausgabeschicht und bewertet so die Qualität des vorliegenden Netzes. Der auf diese Weise bestimmte Fehler wird daraufhin als Funktion der jeweiligen Gewichte des Netzes aufgetragen und anschließend über die Methode des steilsten Abstieges (Gradientenverfahren) versucht, möglichst schnell ein globales Minimum dieser Funktion zu finden. In diesem globalen Minimum ist die Fehlersumme aller Trainingsmuster minimal und das Trainieren des Netzes wird unterbrochen, da eine optimale Beschreibung der Problemstellung gegeben ist. Es ist allerdings von entscheidender Bedeutung ein Übertrainieren des Netzes zu verhindern. Dies wird durch eine parallele Validierung des ANN mit einem von den Trainingsdaten unabhängigen Validierungsdatensatz erreicht.

3. Materialien und Methoden

3.1 Pollenproben

Im Rahmen dieser Arbeit wurden verschiedene Arten von Pollenproben verwendet. Für die Methodenentwicklung dienten kommerziell erhältliche Pollenproben, die von der Firma Sigma Aldrich (Sigma-Aldrich, St. Louis, MO, USA) erworben wurden (Tabelle 3.1). Für die Untersuchung von taxonomischen Verwandtschaftsverhältnissen wurden in den Jahren 2012, 2013 und 2014 Pollenproben in Berliner Parks und dem Botanischen Garten Berlin-Dahlem gesammelt (Tabelle 3.2). Darüber hinaus wurde der Einfluss einer *cherry leaf roll virus* (CLRV) Infektion auf Birkenpollen anhand von Proben untersucht, die in den Jahren 2011 und 2012 an den vier Standorten am Vogelsang (V), Grunewald (GW), im schwarzen Grund (SG) und in der Lentzeallee (LA) gesammelt wurden (Tabelle 3.3). Diese Proben wurden mit *immunocapture reverse transcription-polymerase chain reaction* (IC-RT-PCR) von Kooperationspartnern (Arbeitsgruppe von Prof. Büttner der Lebenswissenschaftlichen Fakultät an der Humboldt-Universität zu Berlin) auf eine CLRV-Infektion getestet. Alle gesammelten Proben wurden in flüssigem Stickstoff schockgefroren und anschließend bei -20°C gelagert. Vor der Untersuchung wurden diese bei Raumtemperatur für 20 Minuten im Exsikkator aufgetaut.

Tabelle 3.1: Die in dieser Arbeit untersuchten, kommerziell erhältlichen Pollenproben

Art	Familie	Ordnung	untersucht mit		
			MALDI-TOF MS	SERS	FTIR
<i>Ambrosia trifida</i>	<i>Asteraceae</i>	<i>Asterales</i>	✓	⊗	⊗
<i>Artemisia absinthium</i>	<i>Asteraceae</i>	<i>Asterales</i>	✓	✓	✓
<i>Artemisia tridentata</i>	<i>Asteraceae</i>	<i>Asterales</i>	✓	✓	✓
<i>Iva xantifolia</i>	<i>Asteraceae</i>	<i>Asterales</i>	⊗	✓	✓
<i>Bassia scoparia</i>	<i>Amaranthaceae</i>	<i>Caryophyllales</i>	⊗	✓	✓
<i>Juniperus scopulorum</i>	<i>Cupressaceae</i>	<i>Coniferales</i>	✓	✓	✓
<i>Betula fontinalis</i>	<i>Betulaceae</i>	<i>Fagales</i>	✓(2 Chargen)	✓	✓
<i>Juglans nigra</i>	<i>Juglandaceae</i>	<i>Fagales</i>	⊗	✓	✓
<i>Populus deltoides</i>	<i>Salicaceae</i>	<i>Malpighiales</i>	✓	✓	✓
<i>Populus nigra</i>	<i>Salicaceae</i>	<i>Malpighiales</i>	✓	✓	✓
<i>Populus nigra 'Italica'</i>	<i>Salicaceae</i>	<i>Malpighiales</i>	✓	⊗	⊗
<i>Populus tremuloides</i>	<i>Salicaceae</i>	<i>Malpighiales</i>	⊗	✓	✓
<i>Populus trichocarpa</i>	<i>Salicaceae</i>	<i>Malpighiales</i>	⊗	✓	✓
<i>Secale cereale</i>	<i>Poaceae</i>	<i>Poales</i>	✓	✓	✓
<i>Sorghum halepense</i>	<i>Poaceae</i>	<i>Poales</i>	⊗	✓	✓
<i>Ulmus pumila</i>	<i>Ulmaceae</i>	<i>Rosales</i>	⊗	✓	✓

Tabelle 3.2: Die in dieser Arbeit untersuchten Pollenproben aus dem Feld

Art	Familie	Ordnung	Anzahl untersuchter Proben mit		
			MALDI-TOF MS	SERS	FTIR
<i>Juniperus communis</i>	<i>Cupressaceae</i>	<i>Coniferales</i>	5	5	4
<i>Juniperus rigida</i>	<i>Cupressaceae</i>	<i>Coniferales</i>	2	2	2
<i>Abies concolor</i>	<i>Pinaceae</i>	<i>Coniferales</i>	4	4	4
<i>Abies pinsapo</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	1
<i>Abies koreana</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	1
<i>Picea abies</i>	<i>Pinaceae</i>	<i>Coniferales</i>	2	2	3
<i>Picea asperata</i>	<i>Pinaceae</i>	<i>Coniferales</i>	6	6	3
<i>Picea breweriana</i>	<i>Pinaceae</i>	<i>Coniferales</i>	2	2	0
<i>Picea chihuahuana</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Picea engelmannii</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Picea glauca</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Picea mariana</i>	<i>Pinaceae</i>	<i>Coniferales</i>	2	2	0
<i>Picea obovata</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Picea omorika</i>	<i>Pinaceae</i>	<i>Coniferales</i>	4	4	3
<i>Picea purpurea</i>	<i>Pinaceae</i>	<i>Coniferales</i>	3	3	0
<i>Picea sitchensis</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Picea spinulosa</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus banksiana</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus bhutanica</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus densiflora</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus heldreichii</i>	<i>Pinaceae</i>	<i>Coniferales</i>	2	2	0
<i>Pinus mugo</i>	<i>Pinaceae</i>	<i>Coniferales</i>	12	12	3
<i>Pinus nigra</i>	<i>Pinaceae</i>	<i>Coniferales</i>	4	4	3
<i>Pinus peuce</i>	<i>Pinaceae</i>	<i>Coniferales</i>	3	3	0
<i>Pinus x rhaetica</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus strobus</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	2
<i>Pinus sylvestris</i>	<i>Pinaceae</i>	<i>Coniferales</i>	11	11	3
<i>Pinus tabulaeformis</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus unicanta</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Pinus wallichiana</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	1	0
<i>Tsuga canadensis</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	0	0
<i>Tsuga diversifolia</i>	<i>Pinaceae</i>	<i>Coniferales</i>	1	0	0
<i>Taxus baccata</i>	<i>Taxaceae</i>	<i>Coniferales</i>	13	13	4
<i>Taxus canadensis</i>	<i>Taxaceae</i>	<i>Coniferales</i>	3	3	2
<i>Taxus brevifolia</i>	<i>Taxaceae</i>	<i>Coniferales</i>	1	1	1
<i>Alnus alnobetula</i>	<i>Betulaceae</i>	<i>Fagales</i>	3	2	1
<i>Alnus cordata</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	0	1
<i>Alnus firma</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	2	1
<i>Alnus glutinosa</i>	<i>Betulaceae</i>	<i>Fagales</i>	4	4	4
<i>Alnus hirsuta</i>	<i>Betulaceae</i>	<i>Fagales</i>	3	3	0
<i>Alnus incana</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	2	2
<i>Alnus japonica</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	0

Art	Familie	Ordnung	Anzahl untersuchter Proben mit		
			MALDI-TOF MS	SERS	FTIR
<i>Alnus mandshurica</i>	<i>Betulaceae</i>	<i>Fagales</i>	3	3	0
<i>Alnus rubra</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	0
<i>Alnus subcordata</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	0
<i>Betula alleghaniensis</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	1	0
<i>Betula coryfolia</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	2	0
<i>Betula insignis</i>	<i>Betulaceae</i>	<i>Fagales</i>	3	2	2
<i>Betula lenta</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	0
<i>Betula luminifera</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	0
<i>Betula middendorffii</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	0
<i>Betula papyrifera</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	1	0
<i>Betula pendula</i>	<i>Betulaceae</i>	<i>Fagales</i>	9	9	7
<i>Betula tauschii</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	1
<i>Betula tatewakiana</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	2	0
<i>Corylus americana</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	0	0
<i>Corylus avellana</i>	<i>Betulaceae</i>	<i>Fagales</i>	11	11	4
<i>Corylus columna</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	1
<i>Corylus cornuta</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	1
<i>Corylus heterophylla</i>	<i>Betulaceae</i>	<i>Fagales</i>	1	1	1
<i>Corylus maxima</i>	<i>Betulaceae</i>	<i>Fagales</i>	3	3	3
<i>Corylus mandshurica</i>	<i>Betulaceae</i>	<i>Fagales</i>	2	2	1
<i>Myrica pensylvanica</i>	<i>Myricaceae</i>	<i>Fagales</i>	4	4	0
<i>Myrica cerifera</i>	<i>Myricaceae</i>	<i>Fagales</i>	2	2	0
<i>Quercus castaneifolia</i>	<i>Fagaceae</i>	<i>Fagales</i>	1	1	0
<i>Quercus ilicifolia</i>	<i>Fagaceae</i>	<i>Fagales</i>	2	2	0
<i>Quercus imbricaria</i>	<i>Fagaceae</i>	<i>Fagales</i>	1	1	0
<i>Quercus libani</i>	<i>Fagaceae</i>	<i>Fagales</i>	2	2	2
<i>Quercus macrocarpa</i>	<i>Fagaceae</i>	<i>Fagales</i>	1	1	0
<i>Quercus palustris</i>	<i>Fagaceae</i>	<i>Fagales</i>	1	1	0
<i>Quercus pubescens</i>	<i>Fagaceae</i>	<i>Fagales</i>	2	2	0
<i>Quercus pyrinaica</i>	<i>Fagaceae</i>	<i>Fagales</i>	3	3	0
<i>Quercus x richterii</i>	<i>Fagaceae</i>	<i>Fagales</i>	2	2	2
<i>Quercus robur</i>	<i>Fagaceae</i>	<i>Fagales</i>	3	3	2
<i>Quercus rubra</i>	<i>Fagaceae</i>	<i>Fagales</i>	2	2	2
Summe:			188	180	77

Tabelle 3.3: Die in dieser Arbeit untersuchten Birkenpollenproben und Ergebnisse der IC-RT-PCR

Bezeichnung	2011		2012		Bemerkungen
	IC-RT-PCR	Anzahl Proben	IC-RT-PCR	Anzahl Proben	
V0	positiv	2	negativ	2	infiziert mit Tabakmosaikvirus (TMV)
V51	positiv	3	positiv	2	
V53	positiv	2	-	0	
V193	positiv	3	negativ	1	
V196	positiv	1	positiv	2	
V197	negativ	2	negativ	1	
V201	-	0	positiv	1	
V226	-	0	negativ	1	
V274	-	0	negativ	2	
V291	-	0	positiv	2	
GW1	-	0	positiv	2	
GW2	-	0	positiv	2	
GW3	negativ	1	-	0	
GW4	negativ	1	-	0	
GW5	negativ	1	-	0	
GW6	-	0	positiv	2	
GW7	negativ	1	negativ	2	
GW8	-	0	negativ	1	
GW9	-	0	negativ	2	
GW10	negativ	1	negativ	1	
GWA	-	0	positiv	1	
GWC	-	0	negativ	2	
SG4	positiv	3	positiv	1	
SG5	negativ	3	-	0	
SG6	positiv	3	positiv	1	
SG7	negativ	3	negativ	1	
SG9	positiv	3	negativ	1	
SG10	positiv	2	negativ	1	
SGA	-	0	negativ	1	
SGB	-	0	positiv	1	
SGC	-	0	positiv	1	
SGE	-	0	negativ	1	
SGF	-	0	negativ	1	
LA5	negativ	3	negativ	1	
LA22	-	0	positiv	3	
LA23	negativ	3	negativ	1	
LA43	-	0	positiv	1	
LA59	-	0	negativ	2	
LA60	-	0	negativ	1	
LA62	-	0	negativ	2	
LA63	-	0	positiv	2	
LA67	-	0	negativ	2	
LA69	-	0	negativ	2	
LA70	-	0	negativ	2	
LA281	-	0	positiv	2	
LA289	-	0	negativ	2	
	Summe:	41	Summe:	62	

3.2 Raman-Apparatur

Sofern nicht anders angegeben, wurden für die SERS- und Raman-Untersuchungen ein LabRam HR800 (Horiba, Bensheim, Deutschland) mit einem BX41 Mikroskop (Olympus, Hamburg, Deutschland) und einem Objektiv (50x) für feste Proben, bzw. einem Wasserimmersionsobjektiv (60x) für flüssige Proben verwendet. Das Anregungslicht mit einer Wellenlänge von 785 nm wurde durch einen Diodenlaser (Toptica, Germany) erzeugt. Das Raman-Licht wurde anschließend, nachdem das Rayleigh-Licht durch einen Kantenfilter unterdrückt wurde, mit einer mit Stickstoff gekühlten CCD-Kamera detektiert (1024×256 Pixel, Horiba).

3.3 Raman-Untersuchungen am Raffinat des Pollenextrakts

Eine Spatelspitze von Pollenkörnern der Art *Betula pendula* wurde 5 Minuten mit einem ml Wasser extrahiert. Anschließend wurde das Wasser abpipettiert, das Raffinat fünfmal mit einem ml Wasser gewaschen und über Nacht bei Raumtemperatur getrocknet. Am nächsten Tag wurde der feste Rückstand auf einen CaF_2 -Träger überführt und jeweils von 3 Pollenkörnern Spektren im Abstand von 2 μl aufgenommen. Dabei wurden jeweils 5 Raman-Spektren, die mit einer Anregungsintensität von $7 \cdot 10^5 \text{ W/cm}^2$ und einer Akkumulationszeit von 5 Sekunden erhalten wurden, gemittelt. Als Referenzmessung wurden jeweils 2 unbehandelte Pollenkörner auf gleiche Weise untersucht.

3.4 SERS-Untersuchungen von Pollenextrakten mit Nanopartikeln in Lösung

Um zu überprüfen, ob SERS in Nanopartikellösung für die Klassifizierung von Pollenproben eingesetzt werden kann, wurden jeweils 2 Pollenextrakte aus 0.3 mg von 8 verschiedenen, kommerziell erhältlichen Pollenproben und 100 μl deionisiertem Wasser hergestellt. Anschließend wurde die Suspension zentrifugiert und der flüssige Überstand von dem festen Raffinat getrennt. 10 μl des Extrakts wurden daraufhin mit 20 μl der Goldnanopartikellösung, die nach [82] hergestellt wurde, und 2 μl NaCl-Lösung (0.1 M) gemischt und 20 μl des Gemisches auf einen CaF_2 -Träger pipettiert. Es wurden jeweils 500 Spektren mit einer Anregungsintensität von $2,9 \cdot 10^5 \text{ W/cm}^2$ und einer Integrationszeit von einer Sekunde aufgenommen und jedes Extrakt zweimal untersucht.

Für die taxonomische Klassifizierung kommerziell erhältlicher Pollenproben wurden jeweils zweimal 0,2 mg der Pollenprobe (siehe Tabelle 3.1) mit 100 μl deionisiertem Wasser 5 Minuten extrahiert. Danach wurde die Suspension zentrifugiert und der flüssige Überstand

von dem festen Raffinat getrennt. Für die SERS-Untersuchungen wurden dann 2 µl des jeweiligen Extraktes mit 2 µl NaCl-Lösung (0.1 M) und 20 µl der Goldnanopartikellösung, die nach [82] hergestellt wurde, gemischt und 20 µl des Gemisches auf einen CaF₂-Träger pipettiert. Es wurden jeweils 500 Spektren mit einer Anregungsintensität von $2,9 \cdot 10^5 \text{ W/cm}^2$ und einer Integrationszeit von einer Sekunde aufgenommen und jedes Extrakt viermal untersucht.

Für die taxonomische Klassifizierung von Pollenproben aus dem Feld wurden jeweils 0,2 mg der Probe (siehe Tabelle 3.2) mit 100 µl deionisiertem Wasser 5 Minuten extrahiert. Danach wurde die Suspension zentrifugiert und der flüssige Überstand von dem festen Raffinat getrennt. Für die SERS-Untersuchungen wurden dann 2 µl des jeweiligen Extraktes mit 2 µl NaCl-Lösung (0.1 M) und 20 µl der Goldnanopartikellösung, die nach [82] hergestellt wurde, gemischt und 20 µl des Gemisches auf einen CaF₂-Träger pipettiert. Es wurden jeweils 500 Spektren mit einer Anregungsintensität von $1,4 \cdot 10^6 \text{ W/cm}^2$ und einer Integrationszeit von einer Sekunde aufgenommen und jedes Extrakt zweimal untersucht.

Für die Identifizierung CLRV-infizierter Birkenpollenproben wurden jeweils 0,2 mg der Probe (siehe Tabelle 3.3) mit 100 µl deionisiertem Wasser 5 Minuten extrahiert. Danach wurde die Suspension zentrifugiert und der flüssige Überstand von dem festen Raffinat getrennt. Für die SERS-Untersuchungen wurden dann 2 µl des jeweiligen Extraktes mit 2 µl NaCl-Lösung (0.1 M) und 20 µl der Goldnanopartikellösung, die nach [82] hergestellt wurde, gemischt und 20 µl des Gemisches auf einen CaF₂-Träger pipettiert. Es wurden jeweils 1000 Spektren mit einer Anregungsintensität von $7,1 \cdot 10^5 \text{ W/cm}^2$ und einer Integrationszeit von einer Sekunde aufgenommen und jedes Extrakt zweimal untersucht. Für die Untersuchung der Pollenproben aus dem Jahr 2011 wurde ein Raman-Mikroskop verwendet, dass mit einem Wasser-immersionsobjektiv (60x, Olympus, Hamburg, Deutschland), einem Diodenlaser (Toptica, Deutschland) und einer mit Stickstoff gekühlten CCD-Kamera (1024 × 256 pixels, Horiba) ausgestattet war.

3.5 MALDI-TOF MS-Untersuchungen

Für die Untersuchungen wurde ein Autoflex III MALDI-TOF Massenspektrometer (Bruker, Daltonik, Deutschland) im linearen Modus mit einem Smartbeam frequenzverdreifachten Nd:Yag-Laser ($\lambda=355\text{nm}$) verwendet. 500 Spektren an 4 verschiedenen Stellen des MALDI-Targets wurden jeweils für 1 Spektrum akkumuliert.

Für die Untersuchung kommerziell erhältlicher Pollenproben wurden jeweils 24 mal etwa 10 mg der Proben (siehe Tabelle 3.1) in 70 µl 70%iger Ameisensäure suspendiert und 10

Sekunden mittels Vortexmischer vermennt. Anschließend wurden jeweils 70 µl Acetonitril hinzugegeben und die Suspensionen erneut für 10 Sekunden mittels Vortexmischer vermennt. Danach wurden die Suspensionen für 30 Sekunden zentrifugiert und jeweils 1 µl des klaren Überstandes auf das MALDI-Target pipettiert. Daraufhin wurde jeweils 1 µl der Matrixlösung (10 mg/ml α -Cyanozimtsäure in 1:1 Wasser:Acetonitril und 0.1 % Trifluoressigsäure) auf die eingetrocknete Probenlösung pipettiert und das MALDI-Target nach erneuter Lösungsmittelverdampfung in das Massenspektrometer überführt. Es wurden Ionen im Bereich m/z 1000 bis 20000 detektiert.

Für die Untersuchung von Pollenproben aus dem Feld und von Birkenpollenproben zur Untersuchung einer CLRV-Infektion wurde jeweils eine Spatelspitze der jeweiligen Probe (siehe Tabelle 3.2 und Tabelle 3.3) auf das MALDI-Target aufgebracht. Anschließend wurde jeweils 1 µl 70%iger Ameisensäure und nach dem Eintrocknen 1 µl der Matrixlösung (10 mg/ml α -Cyanozimtsäure in 1:1 Wasser:Acetonitril und 0.1 % Trifluoressigsäure) auf die Pollen pipettiert und erneut eingetrocknet. Nach dem Überführen des Targets in das Massenspektrometer wurden jeweils 2 Spektren im Bereich zwischen m/z 500 und 15000 aufgenommen.

3.6 FTIR-Untersuchungen einzelner Pollenkörner

Die FTIR-Spektren wurden an der IRIS Beamline des Elektronenspeicherrings Bessy II aufgenommen. Als intensive IR-Lichtquelle diente dabei Synchrotronstrahlung in einem Nicolet Spektrometer (Thermo Scientific, Waltham, USA), das an ein IR-Mikroskop gekoppelt war. Es wurde ein Quecksilber-Cadmium-Tellur-Detektor (*mercury cadmium telluride*, MCT), ein Objektiv (32x) und ein beweglicher Probenstisch zum Untersuchen der Pollenkörner verwendet. Die Spektren wurden in einem Spektralbereich zwischen 1000 und 4000 cm^{-1} unter Verwendung einer Blackmann-Harris 3-Term Apodisierung, eines Zerofillfaktors von 2, einer Mertz-Phasenkorrektur und einer Auflösung von 4 cm^{-1} aufgenommen.

Für die Untersuchung kommerziell erhältlicher Pollenproben (siehe Tabelle 3.1) wurden jeweils 25 Spektren mit einer Aperturgröße von 10 mal 10 µm aufgenommen.

Für die Untersuchung von Pollenproben aus dem Feld wurden jeweils ca. 20 Spektren von 77 Proben von 33 Arten und 9 Gattungen der Ordnungen *Fagales* und *Coniferales* aufgenommen (siehe Tabelle 3.2). Für Pollenproben der Gattungen *Picea* und *Abies*, die durch besonders große Pollenkörner gekennzeichnet sind, wurde dabei eine Aperturgröße von 30 mal 30 µm und für die anderen Pollenproben eine Aperturgröße von 15 mal 15 µm verwendet.

3.7 FTIR-Untersuchungen von Kaliumbromidpresslingen CLRV-infizierter Birkenpollenproben

Die FTIR-Untersuchungen wurden an einem IFS66V (Bruker, Billerica, USA) mit einer Global Lichtquelle, einem Kaliumbromid-Strahlteiler und einem deuterierten-Triglycinsulfat-Detektor (DTGS) durchgeführt. Es wurde jeweils zweimal 1 mg der jeweiligen Birkenpollenprobe (siehe Tabelle 3.3) mit 299 mg Kaliumbromid gemischt. Die Mischungen wurden anschließend in einer Kugelmühle gemahlen und zu Kaliumbromid-Presslingen gepresst. Daraufhin wurden von jedem Pressling jeweils 2 Spektren in einem Spektralbereich zwischen 400 und 4000 cm^{-1} unter Verwendung einer Blackmann-Harris 3-Term Apodisierung, eines Zerofillfaktors von 2, einer Mertz-Phasenkorrektur und einer Auflösung von 4 cm^{-1} aufgenommen.

3.8 ESEM Untersuchungen von Pollenkörnern

Die ESEM Untersuchungen wurden an einem Hitachi TM-1000 und einem Sekundärelektronendetektor bei einer Beschleunigungsspannung von 15 kV durchgeführt.

3.9 Vorbehandlung und Auswertung von Pollendaten

Sämtliche Spektrenvorbehandlungen und multivariaten Auswertungen, sowie die Transformation der Daten zu einer Standardnormalverteilung vor der Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) wurden unter Verwendung der Software Matlab (The Mathworks, Inc., Natick, MA, USA) durchgeführt. Bei der hierarchischen Clusteranalyse (HCA) wurden euklidische Distanzen und der Ward's Algorithmus verwendet.

3.10 Spektrenvorbehandlung und Auswertung der SERS-Spektren durch HCA und PCA

Alle Raman- und SERS-Spektren wurden jeweils anhand eines Spektrums von 4-Acetamidophenol frequenzkalibriert, im Bereich zwischen 400 und 1700 cm^{-1} interpoliert und vektornormiert. Für den Vergleich von Mittelwertspektren wurden die vektornormierten Spektren der Pollenkörner des Raffinats und der unbehandelten Pollenkörner, sowie jeweils die 500 vektornormierten SERS-Spektren kommerziell erhältlicher Pollen gemittelt. Es wurden erste Ableitungen der vektornormierten Spektren gebildet und bei allen SERS-

Datensätzen die 500, bzw. 1000 Spektren einer Probe gemittelt. Anschließend wurden die Spektren mit HCA und PCA ausgewertet.

3.11 Klassifizierung von SERS-Spektren durch künstliche neuronale Netze

Für die Auswertung der SERS-Spektren kommerziell erhältlicher Pollenproben durch ein künstliches neuronales Netz (*artificial neural network*, ANN) wurden jeweils die Daten eines Extraktes (28000 Spektren) für das Aufbauen und die Daten des anderen Extraktes (28000 Spektren) für das Testen des Netzes verwendet. Die Daten, die zum Aufbauen des Netzes verwendet wurden, wurden weiterhin unterteilt in Trainings- (70%), Validierungs- (25%) und Testdaten (5%). Ein Netzwerk aus einer Eingangsschicht von 894 Neuronen, einer versteckten Schicht von 50 Neuronen und einer Ausgangsschicht von 14 Neuronen wurde erstellt und mit conjugate gradient backpropagation trainiert. Das Training wurde beim minimalen Validierungsfehler, der nach 356 Iterationen erreicht war, beendet. Anschließend wurden die Testdaten benutzt, um das ANN zu bewerten. Die Zuordnungen der jeweils 500 Spektren wurden gemittelt und die jeweiligen Datensätze durch die winner-takes-all (WTA) Methode^[35] klassifiziert.

Auch die Daten der Pollenproben aus dem Feld wurden in einen Trainings- und einen reinen Testdatensatz geteilt. Das ANN wurde aus einer Eingangsschicht von 894 Neuronen, einer versteckten Schicht von 50 Neuronen und einer Ausgangsschicht von 14 Neuronen erstellt und mit conjugate gradient backpropagation Algorithmus trainiert. Das Trainieren der ANN und die Klassifizierung wurden jeweils separat für die verschiedenen taxonomischen Gruppen durchgeführt. Dabei wurde jeweils der Trainingsdatensatz in 70% Trainings- 25% Validierungs- und 5% Testdatensatz geteilt und der reine Testdatensatz ausschließlich dazu verwendet, um das neuronale Netz zu testen. Die Zuordnungen der jeweils 500 Spektren des Testdatensatzes wurden gemittelt und die jeweiligen Datensätze durch die WTA-Methode klassifiziert.

Für die Klassifizierung auf taxonomischer Ebene der Ordnung bestand der Trainingsdatensatz aus 40000 Spektren von 40 Proben der Ordnung *Coniferales* und 44000 Spektren von 44 Proben der Ordnung *Fagales*. Der minimale Validierungsfehler wurde nach 407 Iterationen erreicht und der Testdatensatz bestehend aus 55000 Spektren von 55 Proben der Ordnung *Coniferales* und 41000 Spektren von 41 Proben der Ordnung *Fagales* wurde verwendet, um das Netz zu testen.

Für die Klassifizierung auf taxonomischer Ebene der Gattung wurde jeweils ein Netz für die Spektren der beiden Ordnungen trainiert. Der Trainingsdatensatz der Ordnung *Coniferales*

bestand aus 9000 Spektren von 9 Proben der Gattung *Taxus*, jeweils 12000 Spektren von 12 Proben der Gattungen *Picea* und *Pinus*, 4000 Spektren von 4 Proben der Gattung *Juniperus* und 3000 Spektren von 3 Proben der Gattung *Abies*. Der minimale Validierungsfehler wurde nach 231 Iterationen erreicht und der Testdatensatz bestehend aus 8000 Spektren von 8 Proben der Gattung *Taxus*, 13000 Spektren von 13 Proben der Gattung *Picea*, jeweils 3000 Spektren von 3 Proben der Gattungen *Juniperus* und *Abies*, sowie 28000 Spektren von 28 Proben der Gattung *Pinus* wurde verwendet, um das Netz zu testen.

Der Trainingsdatensatz der Ordnung *Fagales* bestand aus 11000 Spektren von 11 Proben der Gattung *Betula*, jeweils 10000 Spektren von 10 Proben der Gattungen *Alnus*, *Quercus* und *Corylus*, sowie 3000 Spektren von 3 Proben der Gattung *Myrica*. Der minimale Validierungsfehler wurde nach 301 Iterationen erreicht und der Testdatensatz bestehend aus jeweils 10000 Spektren von 10 Proben der Gattungen *Betula* und *Quercus*, jeweils 9000 Spektren von 9 Proben der Gattungen *Alnus* und *Corylus*, sowie 3000 Spektren von 3 Proben der Gattung *Myrica* wurde verwendet, um das Netz zu testen.

Für die Klassifizierung auf taxonomischer Ebene der Art wurde ein Testdatensatz aus jeweils 5000 Spektren von 5 Proben der Art *Pinus mugo* und *Pinus sylvestris* verwendet. Der minimale Validierungsfehler wurde nach 34 Iterationen erreicht und der Testdatensatz bestehend aus 7000 Spektren von 7 Proben der Art *Pinus mugo* und 6000 Spektren von 6 Proben der Art *Pinus sylvestris* wurde verwendet, um das Netz zu testen.

Auch die SERS-Spektren von Birkenpollenproben zur Identifizierung einer CLRV-Infektion wurden in einen Trainings- und einen reinen Testdatensatz geteilt. Das ANN wurde aus einer Eingangsschicht von 894 Neuronen, einer versteckten Schicht von 50 Neuronen und einer Ausgangsschicht von 2 Neuronen erstellt und mit conjugate gradient backpropagation Algorithmus trainiert. Der Trainingsdatensatz wurde erneut in 70% Trainings- und 30% Validierungsdatensatz geteilt und der reine Testdatensatz nach dem Training des Netzes dazu verwendet, um das neuronale Netz zu testen. Die Zuordnungen der jeweils 1000 Spektren des Testdatensatzes wurden gemittelt und die jeweiligen Datensätze durch die WTA-Methode klassifiziert.

Für die Klassifizierung der Pollenproben aus dem Jahr 2011 bestand der Trainingsdatensatz für die Klassen der CLRV-positiv und der CLRV-negativ getesteten Birkenpollen jeweils aus 18000 Spektren von 9 Proben, die von jeweils 4 verschiedenen Individuen gesammelt wurden. Der minimale Validierungsfehler wurde nach 23 Iterationen erreicht und der Testdatensatz bestehend aus 46000 Spektren von 23 Proben, die von 11 verschiedenen Individuen gesammelt wurden, wurde verwendet, um das Netz zu testen.

Für die Klassifizierung der Pollenproben aus dem Jahr 2012 bestand der Trainingsdatensatz aus 20000 Spektren von 10 Proben, die von jeweils 6 verschiedenen Individuen gesammelt wurden. Der minimale Validierungsfehler wurde nach 236 Iterationen erreicht und der Testdatensatz bestehend aus 84000 Spektren von 42 Proben, die von 34 verschiedenen Individuen gesammelt wurden, wurde verwendet, um das Netz zu testen.

3.12 Extraktion von Spektren aus SERS-Daten von Pollenproben aus dem Feld

Die für die Klassifizierung verwendeten Spektren der Pollenextrakte von Pollenproben aus dem Feld (siehe Absatz 3.4) wurden auch dazu verwendet, um neuronale Netze zu erhalten, die für die Extraktion von Spektren dienten. Es wurden dazu ANN aus einer Eingangsschicht von 894 Neuronen, einer versteckten Schicht von 50 Neuronen und einer Ausgangsschicht von 14 Neuronen erstellt und mit conjugate gradient backpropagation Algorithmus trainiert. Das Trainieren des ANN und die anschließende Extraktion wurde separat für die verschiedenen taxonomischen Gruppen durchgeführt und dabei jeweils der Trainingsdatensatz in 70% Trainings- und 30% Validierungsdatensatz geteilt. Die Spektren der jeweiligen Probe, die durch das ANN extrahiert wurden, wurden anschließend gemittelt und die gemittelten Spektren mit multivariaten Verfahren ausgewertet.

Für die Extraktion von Spektren für die Klassifizierung auf taxonomischer Ebene der Ordnung bestand der Trainingsdatensatz aus 46000 Spektren von 46 Proben der Ordnung *Coniferales* und 47000 Spektren von 47 Proben der Ordnung *Fagales*. Der minimale Validierungsfehler wurde nach 302 Iterationen erreicht. Anschließend wurden jeweils alle Spektren beider Ordnungen durch das ANN getestet und die Spektren als taxonomisch spezifisch definiert und extrahiert, die eine Zuordnung zur jeweiligen Ordnung aufwiesen, die höher als 0.9 war. Zusätzlich dazu wurden die Spektren als taxonomisch relevant definiert, die jeweils eine Zuordnung zu einer der Ordnungen aufwiesen, die höher als 0.9 war.

Für die Extraktion von Spektren für die Klassifizierung auf taxonomischer Ebene der Gattung wurde jeweils ein Netz für die Spektren der beiden Ordnungen trainiert. Der Trainingsdatensatz der Ordnung *Coniferales* bestand aus 9000 Spektren von neun Proben der Gattung *Taxus*, jeweils 12000 Spektren von zwölf Proben der Gattungen *Picea* und *Pinus*, 7000 Spektren von sieben Proben der Gattung *Juniperus* und 6000 Spektren von sechs Proben der Gattung *Abies*. Der Trainingsdatensatz der Ordnung *Fagales* bestand aus 11000 Spektren von elf Proben der Gattung *Betula*, jeweils 10000 Spektren von zehn Proben der Gattungen *Alnus*, *Quercus* und *Corylus*, sowie 6000 Spektren von sechs Proben der Gattung *Myrica*. Der minimale Validierungsfehler wurde nach 288, bzw. 194 Iterationen erreicht. Anschließend

wurden jeweils alle Spektren der verschiedenen Gattungen durch das jeweilige ANN getestet und die Spektren als taxonomisch spezifisch definiert und extrahiert, die eine Zuordnung zur jeweiligen Gattung aufwiesen, die höher als 0.5 war. Zusätzlich dazu wurden die Spektren als taxonomisch relevant definiert, die jeweils eine Zuordnung zu irgendeiner Gattung aufwiesen, die höher als 0.5 war.

3.13 Spektrenvorbehandlung und multivariate Auswertung der MALDI-TOF MS-Spektren

Die MALDI-TOF MS-Spektren wurden anhand der charakteristischen Massen von den Biopolymerstandards Insulin, Cytochrom C, Myoglobin und Ubiquitin kalibriert. Anschließend wurden die Rohdaten der kommerziell erhältlichen Pollenproben im Massenbereich zwischen m/z 1001 bis 15001 mit einer Schrittweite von 2 interpoliert, vektornormiert, mit einem Polynom sechsten Grades basislinienkorrigiert und eine HCA durchgeführt. Die Rohdaten der Pollenproben aus dem Feld wurden im Massenbereich zwischen m/z 1000 bis 10000 und die der Birkenpollenproben zur Untersuchung der CLRV-Infektion zwischen m/z 4400 bis 8000 mit einer Schrittweite von 2 interpoliert und mit einem Polynom sechsten Grades basislinienkorrigiert. Danach wurden jeweils Hauptkomponentenanalysen nach Vektornormierung für alle Datensätze und hierarchische Clusteranalysen nach Normierung auf den jeweiligen maximalen Peak für die Datensätze der Pollenproben aus dem Feld durchgeführt.

3.14 Klassifizierung von FTIR-Spektren einzelner Pollenkörner

Es wurden zunächst Spektren, die in der PCA eine sehr große Varianz zu den anderen Spektren aufwiesen, als Ausreiser identifiziert und aus den spektralen Datensätzen entfernt. Um die verwendete Anzahl an Savitzky-Golay (SG) Datenpunkten zu optimieren, wurden anschließend die Daten des Fingerprintbereichs von 1200 bis 2000 cm^{-1} extrahiert, jeweils mit der variierenden Anzahl an SG-Datenpunkten von 3 bis 81 die zweite Ableitung gebildet und eine EMSC-Korrektur der Spektren durchgeführt. Aus den auf diese Weise erhaltenen Datensätzen wurde für jede verwendete Anzahl an SG-Datenpunkten ein PLSR-Modell aufgebaut, bei dem die Parameter der korrekt klassifizierten Spektren, der verwendeten PLSR-Komponenten und des Bestimmtheitsmaßes R^2 , bestimmt wurden. Die Anzahl der verwendeten PLSR-Komponenten und das Bestimmtheitsmaßes R^2 ergab sich dabei aus einem χ^2 -Test, bzw. einer Kreuzvalidierung der Daten.

Für die Auswertung kommerziell erhältlicher Pollenproben wurde ein hierarchisches Klassifizierungsmodell aufgebaut, bei dem die Optimierung der Anzahl an SG-Datenpunkten durch die Verwendung aller Spektren erfolgte. Anschließend wurden jeweils 10 Spektren jeder Pollenart für das Trainieren und die jeweils übrigen Spektren für das Testen des hierarchischen PLSR-Modells verwendet.

Für die Auswertung von Pollenproben aus dem Feld wurde anhand der taxonomischen Verwandtschaftsverhältnisse ein hierarchisches Klassifizierungsmodell aufgebaut, bei dem die Optimierung der Anzahl an SG-Datenpunkten durch Verwenden aller Spektren erfolgte. Anschließend wurden die Spektren in einen Trainingsdatensatz aus insgesamt 784 Spektren von 39 Proben und einen Testdatensatz aus insgesamt 768 Spektren von 38 Proben, geteilt. Der Trainingsdatensatz bestand dabei aus 103 Spektren von 5 Proben der Gattung *Betula*, 116 Spektren von 6 Proben der Gattung *Corylus*, 80 Spektren von 4 Proben der Gattung *Alnus*, 82 Spektren von 4 Proben der Gattung *Quercus*, 59 Spektren von 3 Proben der Gattung *Juniperus*, 80 Spektren von 4 Proben der Gattung *Taxus*, 81 Spektren von 4 Proben der Gattung *Picea*, 61 Spektren von 3 Proben der Gattung *Abies* und 122 Spektren von 6 Proben der Gattung *Pinus*. Der Testdatensatz bestand aus 102 Spektren von 5 Proben der Gattung *Betula*, 99 Spektren von 5 Proben der Gattung *Corylus*, 105 Spektren von 5 Proben der Gattung *Alnus*, 80 Spektren von 4 Proben der Gattung *Quercus*, 61 Spektren von 3 Proben der Gattung *Juniperus*, 57 Spektren von 3 Proben der Gattung *Taxus*, 102 Spektren von 5 Proben der Gattung *Picea*, 61 Spektren von 3 Proben der Gattung *Abies* und 101 Spektren von 5 Proben der Gattung *Pinus*.

3.15 Spektrenvorbehandlung und multivariate Auswertung von FTIR-Spektren von Kaliumbromidpresslingen

Die Spektren wurden vektornormiert, abgeleitet und anschließend mit einer PCA ausgewertet.

3.16 Spektrenvorbehandlung und multivariate Auswertung von Pollendaten durch CPCA

Die *Consensus Principal Component Analysis* (CPCA) wurde angewendet, um die Daten der einzelnen Methoden zu kombinieren.

Von den frequenzkalibrierten, interpolierten und vektornormierten SERS-Spektren wurde jeweils die erste Ableitung gebildet und ausgewählte Spektren mit ANN extrahiert (siehe Abschnitt 3.12). Diese wurden gemittelt und zu einer Standardnormalverteilung transformiert.

Darüber hinaus wurden bei der Klassifizierung auf Ebene der taxonomischen Ordnung auch gemittelte SERS-Spektren verwendet, die nicht zuvor durch das ANN selektiert wurden.

Die MALDI-TOF MS-Spektren wurden interpoliert, vektornormiert und basislinienkorrigiert.

Die FTIR-Spektren der Kaliumbromidpresslinge wurden vektornormiert und es wurde die erste Ableitung gebildet.

4. SERS zur taxonomischen Klassifizierung von Pollen

Die oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS) bietet zahlreiche Vorteile für die Charakterisierung biologischer Proben. Aufgrund der hohen lokalen Verstärkung der Raman-Signale können dabei biophysikalische Prozesse in vivo unter biologisch normalen Bedingungen mit hoher räumlicher und zeitlicher Auflösung untersucht werden.^[57] Dies ist u.a. bei der Untersuchung von Zellen vorteilhaft. So wurde bereits Anfang der 90er Jahre SERS dazu verwendet, um Cytostatika in lebenden Tumorzellen zu lokalisieren.^[83] Auch die Verteilung von charakteristischen Zellmolekülen, wie z.B. DNA oder Phenylalanin innerhalb einzelner, lebender Zellen kann auf diese Weise analysiert werden.^{[84],[85]} Bei Anwendungen wie diesen zeigt sich, dass SERS als komplementäre Methode zur weit verbreiteten Fluoreszenzspektroskopie dienen und dabei den Vorteil einer markierungsfreien Detektion bieten kann. Dies zeigt besonders die Tatsache, dass mit SERS ebenfalls einzelne Moleküle untersucht werden können.^{[86],[87]} Es ist jedoch auch möglich SERS-Substrate mit spezifischen Reportern zu koppeln, um gezielt Eigenschaften von biologischen Proben zu untersuchen. So kann z.B. der pH-Wert lebender Zellen mit SERS von 4-Mercaptobenzoessäure^[88] oder der Glukosegehalt des Blutes bestimmt werden^[89]. Durch die Verwendung verschiedener Reportermoleküle können auch bildgebende Verfahren entwickelt werden, durch die gleichzeitig untersucht werden kann, wo sich verschiedene Moleküle in einer Zelle befinden.^[90] Hier zeigte sich, dass auch SERS-Daten erfolgreich mit multivariaten Methoden ausgewertet werden können. Aufgrund der „vermeintlich“ geringen Reproduzierbarkeit der SERS-Untersuchung ist dies bisher jedoch relativ selten realisiert worden. Es konnten allerdings Methoden entwickelt werden, um Tumore in Blutproben zu identifizieren^[91], Krebszellen von gesunden Zellen zu unterscheiden^[92] und Bakterien^[59] und Viren^{[58],[93]} zu klassifizieren. Da mit SERS in relativ kurzer Zeit eine hohe Anzahl an Spektren erhalten werden kann, bieten sich besonders selbstlernende Mustererkennungsverfahren zur Auswertung der Daten an. Dies konnte bereits in ersten Anwendungen von support vector machines (SVM) zur Identifizierung von Tumoren im menschlichen Serum^[94] und zur Klassifizierung von Viren^[95] gezeigt werden. Da künstliche neuronale Netze (ANN) besonders gut geeignet sind, um eine Klassifizierung mehrerer Gruppen zu erreichen, wurden diese auf die Untersuchung von Mischungen verschiedener Biomoleküle zur parallelen Identifizierung einzelner Bestandteile angewendet.^[81] Dabei konnte eindrucksvoll das Potenzial dieser Kombination belegt werden.

Das Ziel dieser Arbeit ist es, reproduzierbare SERS-Daten von wässrigen Pollenextrakten zu erhalten und diese anschließend mit multivariaten Verfahren zur taxonomischen Klassifizierung zu kombinieren. Dass mit SERS die isolierte Pollenhülle analysiert werden kann, wurde bereits demonstriert.^[20] Hier soll allerdings, auch im Gegensatz zu anderen, bisherigen schwingungsspektroskopischen Untersuchungen^{[13],[16]}, bei denen primär die äußere Pollenhülle untersucht wurde, der Fokus auf den inneren, wasserlöslichen Pollenbestandteilen liegen. Es soll dabei zunächst gezeigt werden, dass durch SERS von wässrigen Extrakten tatsächlich die inneren Bestandteile von Pollenkörnern untersucht werden können (Abschnitt 4.1). Anschließend soll eine Methode entwickelt werden, um SERS-Spektren, die reproduzierbaren, taxonomisch spezifischen Fingerabdrücken der Pollen entsprechen sollen, in Nanopartikellösung zu erhalten und auswerten zu können (Abschnitt 4.2). Daraufhin soll die multivariate Auswertung der SERS-Daten optimiert und dabei sowohl unüberwachte, als auch überwachte, multivariate Methoden erprobt werden (Abschnitt 4.3). Die in den ersten Abschnitten gewonnen Erkenntnisse aus der Untersuchung kommerziell erhältlicher Pollenproben sollen daraufhin auf die taxonomische Klassifizierung von Pollen aus dem Feld angewendet werden. Dabei soll überprüft werden, ob sich taxonomische Verwandtschaftsverhältnisse aus den SERS-Daten reproduzieren lassen (Abschnitt 4.4) und, ob klassifizierungsrelevante SERS-Spektren aus den großen Datenmengen extrahiert werden können (Abschnitt 4.5).

4.1 SERS von Pollenextrakten liefert komplementäre schwingungsspektroskopische Informationen

Um SERS sinnvoll für die Klassifizierung von Pollenprobenextrakten einzusetzen, ist es notwendig aufzuklären welche Bestandteile der komplexen Pollenchemie mit dieser Methode untersucht werden können. Dazu sollen hier zunächst Spektren des wässrigen Pollenextraktes qualitativ diskutiert (Abschnitt 4.1.1) und anschließend mit Raman-Spektren verglichen werden (Abschnitt 4.1.2).

4.1.1 Qualitative Diskussion von SERS-Spektren wässriger Pollenextrakte

Bei der Untersuchung eines einzelnen Pollenextraktes wird eine Vielzahl unterschiedlicher SERS-Spektren erhalten. Dies ist darauf zurückzuführen, dass sich jeweils zum Zeitpunkt der Messung unterschiedliche Moleküle und unterschiedliche Nanopartikel, bzw. Nanoaggregate im Fokusvolumen befinden. In Abbildung 4.1 sind repräsentative SERS-Spektren des Extraktes von *Secale cereale*-Pollen gezeigt, um die Variation der Spektren zu demonstrieren.

Die markantesten Banden sind beschriftet und durch den Vergleich mit Referenzspektren wurde eine mögliche Zuordnung durchgeführt (siehe Tabelle 4.1). Die Bande bei 734 cm^{-1} tritt in jedem Spektrum auf und kann der Ringatmungsschwingung von Adenin^{[98],[99],[21]} oder Adeninderivaten, wie z.B: Adenosinphosphaten^{[100],[101]} zugeordnet werden. Ein Grund für die Omnipräsenz dieser Bande ist der hohe Streuquerschnitt dieser Schwingung im Vergleich zu denen anderer funktioneller Gruppen. Der dominierende Einfluss dieser Bande könnte auch durch eine bevorzugte Adsorption an die Goldnanopartikel erklärt werden. Dafür spricht, dass bei der Verwendung von sehr hoch konzentrierten Pollenextrakten fast ausschließlich Banden auftreten, die auf Adenin und Adeninderivate zurückgeführt werden können.^[21]

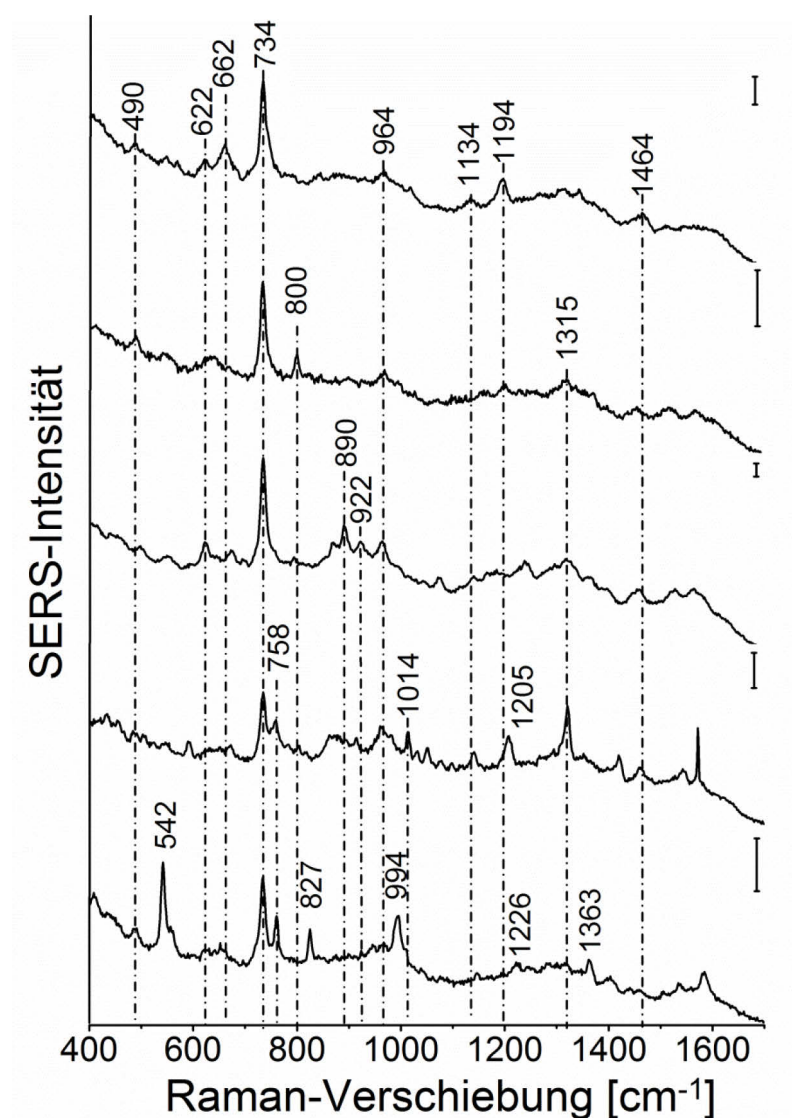


Abbildung 4.1: Repräsentative SERS-Spektren des wässrigen Pollenextraktes von *Secale cereale*-Pollen in Goldnanopartikellösung. Die Maßstabsskala entspricht 500 cps. Siehe Tabelle 4.1 für eine mögliche Zuordnung der charakteristischen Banden.

Tabelle 4.1: Raman-Verschiebungen charakteristischer Banden in den SERS-Spektren und deren mögliche Zuordnung zu Referenzspektren

Raman-Verschiebung [cm^{-1}]	mögliche Zuordnung ^a
1464	G,A: δ, ν^b
1363	Trp: Ring ^b
1315	C: ν^b
1205	His: δ Ring ^c
1194	C: ν^b
1134	G: δ^b
1014	Phe: Ringatmung ^b
994	Trp: asym Indolatmung ^b
964	A,G: δ^b
922	T: δ^b
890	Indol ^d
827	Val: ν^c
800	C: Ringatmung ^b
758	Trp:sym Indolatmung ^b
734	A: Ringatmung ^b
662	G: Ringatmung ^b
622	A: δ^b
542	Indol ^d
490	C,G,T: δ^b

^a Abkürzungen: ν , Streckschwingung; δ , Deformationsschwingung;
A, Adenin; C, Cytosin; T, Thymin; G, Guanin; His, Histidin;
Phe, Phenylalanin; Trp, Tryptophan; Val, Valin

^b nach [21]

^c nach [96]

^d nach [97]

Auch die meisten anderen Banden in den Spektren können Schwingungen von Aminosäuren oder DNA-Basen, die sich vor allem in den inneren Bestandteilen der Pollen befinden, zugeordnet werden. Da mit Raman-^[13] und FTIR-Untersuchungen^[16] primär die Sporopollenin-Außenhülle untersucht wird, liefern SERS-Spektren von Pollenextrakten somit komplementäre schwingungsspektroskopische Informationen über die Pollenbiochemie.

4.1.2 Raman-Untersuchung des Raffinats

Um zu validieren, ob durch die SERS-Experimente tatsächlich andere Bestandteile der komplexen Pollenchemie analysiert werden können, wurden chemische Karten durch Raman-Messungen von intakten, unbehandelten Pollenkörnern und von Pollenkörnern, die sich im festen Rückstand der Extraktion (Raffinat) befanden, aufgenommen und verglichen. Da hier

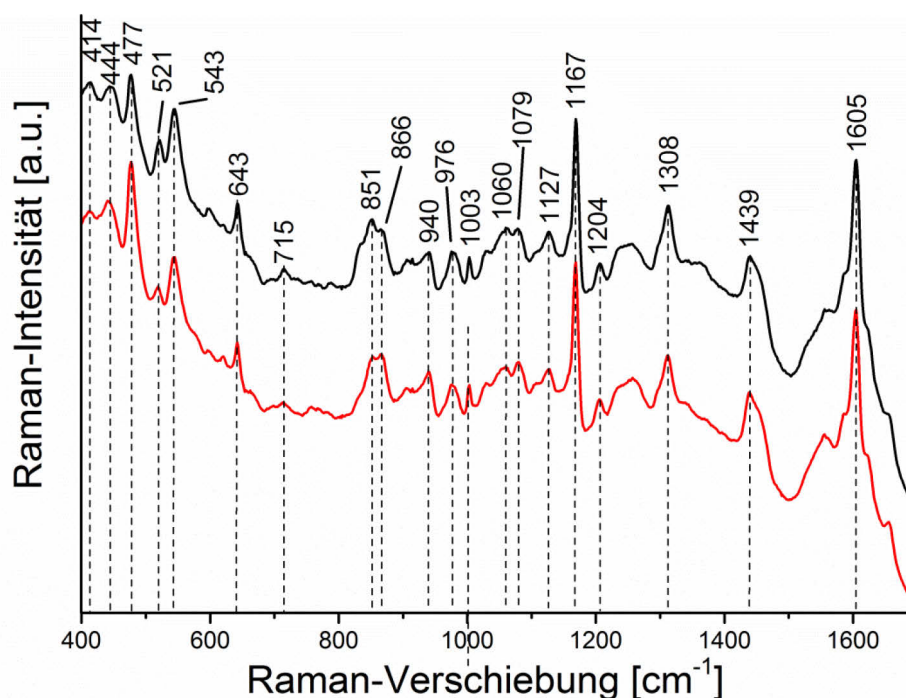


Abbildung 4.2: Mittelwertspektren von intakten, unbehandelten Birkenpollenkörnern (schwarz) und von Pollenkörnern des Raffinats (rot).

Spektren von chemischen Kartierungen ganzer Pollenkörner verwendet wurden, kann dabei untersucht werden, ob lokale Differenzen zwischen den Spektren größer sind als solche, die aus der wässrigen Extraktion resultieren. Die Mittelwerte von Spektren intakter Pollenkörner und Pollenkörner des Raffinats sind in Abbildung 4.2 und eine Zuordnung der intensivsten Banden in Tabelle 4.2 gezeigt. In den Spektren treten jeweils die gleichen Banden auf, was darauf schließen lässt, dass durch die Extraktion die Bestandteile, die in den Raman-Spektren repräsentiert sind, weitestgehend unverändert bleiben. Die Bande bei 477 cm⁻¹ ist im Spektrum von Pollenkörnern des Raffinats jedoch eindeutig intensiver ausgeprägt. Diese Bande kann Gerüstschwingungen von Kohlenhydraten, wie z.B. Pektinen, Stärke oder Zuckern zugeordnet werden (Tabelle 4.2). Um die Varianzen der Daten von unbehandelten Pollenkörnern und Pollenkörnern des Raffinats genauer zu untersuchen, wurde anschließend eine PCA durchgeführt. Eine Unterscheidung zwischen den verschiedenen Pollenkörnern ist anhand der ersten beiden Hauptkomponenten (Abbildung 4.3a) nicht möglich. Betrachtet man die Loadings dieser Hauptkomponenten (Abbildung 4.3c, erste und zweite Spur), so ist erkennbar, dass die Varianzen vor allem auf Unterschiede in Banden bei 478 cm⁻¹, 1169 cm⁻¹ und 1604 cm⁻¹ zurückgeführt werden können. Diese Banden lassen sich Kohlenhydraten (478 cm⁻¹), bzw. charakteristischen Sporopolleninschwingungen (1168 cm⁻¹ und 1604 cm⁻¹) der Außenhülle zuordnen (siehe Tabelle 4.2). Eine Trennung der Spektren, die von unbehandelten Pollenkörnern und von Pollenkörnern des Raffinats erhalten werden, ist anhand der Scores

Tabelle 4.2: Raman-Verschiebungen charakteristischer Banden in den Raman-Spektren und deren mögliche Zuordnung zu Referenzspektren

Raman-Verschiebung [cm^{-1}]	mögliche Zuordnung ^a
1605	Phe ^{b,d} , Tyr ^{b,d} , Trp ^{b,d} , Sporopollenin ^c
1439	Lipide ^{b,d} , Proteine ^{b,d} , Sporopollenin ^c : δ
1308	Nukleotide ^d
1256	Proteine: Amid III ^d
1204	Tyr ^{b,d} , Phe ^{b,d} , Sporopollenin ^c
1167	Carotinoide ^{b,d} , Sporopollenin ^c
1127	Cellulose ^{c,d}
1079	Nukleotide: ν ^{b,d}
1060	Carotinoide: ν ^{b,d}
1003	Phe: Ringatmung ^{b,d}
976	Carotinoide ^{b,d}
940	Proteine ^d
851	Proteine: Tyr: δ ^b
715	Proteine: Met: ν ^b
643	Tyr: ν ^{b,d}
543	Proteine: Cys: ν ^d
521	Proteine: Cys: ν ^{b,d}
477	Kohlenhydrate: δ ^{b,d}
444	Kohlenhydrate: δ ^d
414	Tyr ^d

^a Abkürzungen: ν , Streckschwingung; δ , Deformationsschwingung; Cys, Cystein; Met, Methionin; Phe, Phenylalanin; Trp, Tryptophan; Tyr, Tyrosin

^b nach [13]

^c nach [16]

^d nach [102]

der ersten und zweiten Hauptkomponenten (Abbildung 4.3a) nicht möglich. Wie die Unterschiede im Aufbau und der relativen Konzentration der Bestandteile der Pollenhülle, ist auch die Varianz im Beitrag der Kohlenhydrate (478 cm^{-1}) somit lediglich auf lokale Heterogenität der Pollenkörner und nicht auf die Effekte der Extraktion zurückzuführen.

Die Scores der dritten Hauptkomponente (Abbildung 4.3b) zeigen, dass sich intakte Pollenkörner und Pollenkörner des Raffinats anhand dieser größtenteils unterscheiden. Die Loadings von Hauptkomponente 3 (Abbildung 4.3c, dritte Spur) zeigen, dass eine Vielzahl von Banden zu dieser Unterscheidung beitragen. Diese können Aminosäuren und Proteinen, sowie Nukleotiden zugeordnet werden (siehe Tabelle 4.2). Es lässt sich daraus schlussfolgern, dass sich durch die Extraktion v.a. die Protein- und Nukleotidzusammensetzung der inneren, zellulären Bestandteilen der Pollenkörner verändert. Vermutlich werden diese Bestandteile

aus dem Polleninneren extrahiert, was auch durch die SERS-Spektren der Extrakte (siehe Abschnitt 4.1.1) bestätigt wird. Die dritte Hauptkomponente repräsentiert jedoch nur 5% der gesamten Varianz, während die ersten beiden Hauptkomponenten über 70% der Varianz im Datensatz entsprechen. Daraus folgt, dass man in den Raman-Spektren vor allem die Struktur und Zusammensetzung der Pollenhülle untersuchen kann und dass SERS-Spektren sehr gut dafür geeignet sind, komplementäre Informationen aus dem Polleninneren zu erhalten.

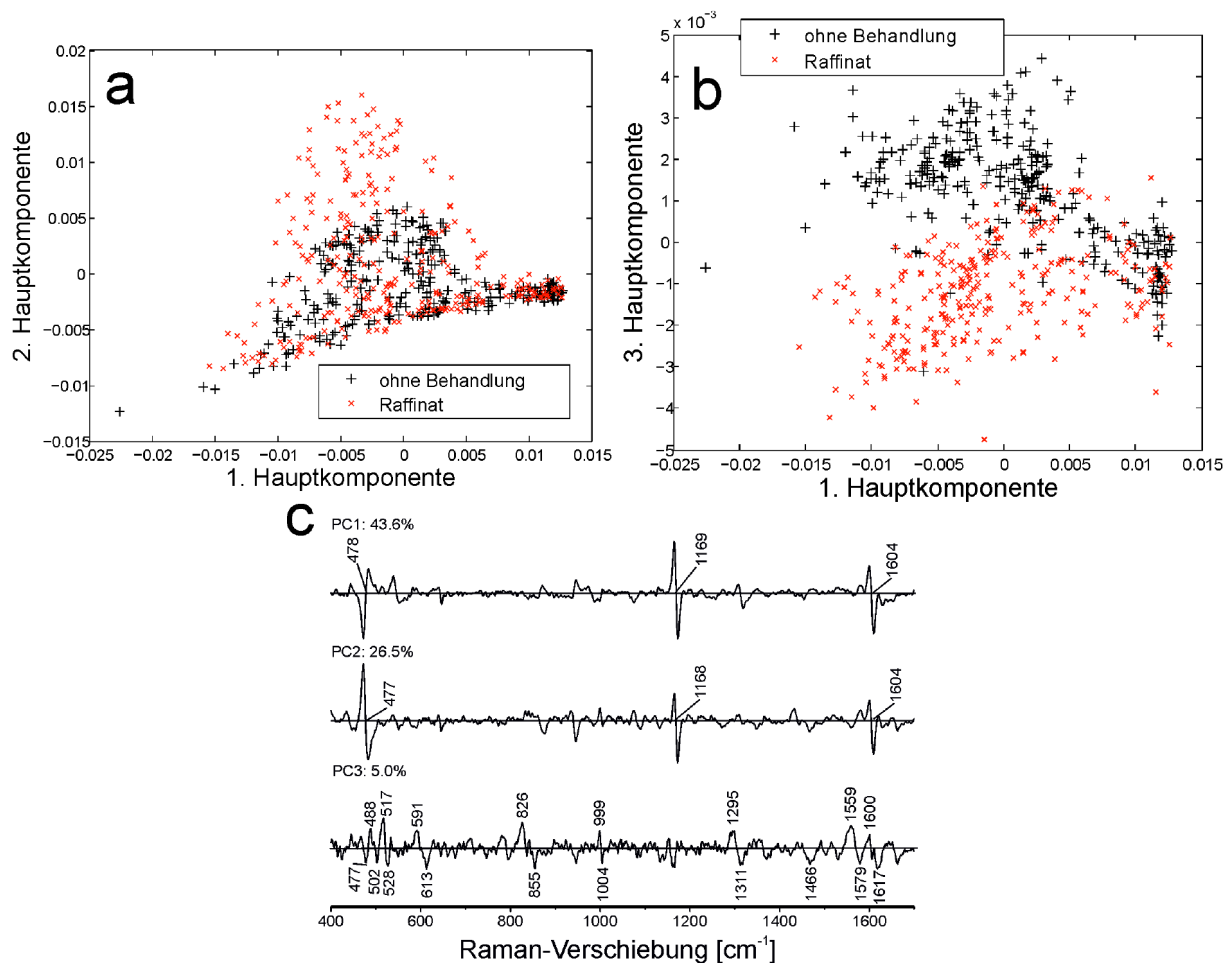


Abbildung 4.3: Ergebnis der PCA von Spektren von unbehandelten Birkenpollenkörnern und von Birkenpollenkörnern des Raffinats. Es sind die Scores der ersten und zweiten Hauptkomponente (a), der ersten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der entsprechenden Hauptkomponenten (c) gezeigt.

4.2 Eignung von SERS in Nanopartikellösung zur Klassifizierung von Pollenproben

In Abbildung 4.1 wird deutlich, dass SERS-Spektren, die in Nanopartikellösung erhalten werden im Vergleich zu FTIR- und Raman-Spektren eine geringe Reproduzierbarkeit aufweisen. Diese wird seit mehreren Jahrzehnten in Bezug auf mögliche analytische Anwendungen von SERS kontrovers diskutiert. Es wurde bereits gezeigt, dass es möglich ist durch Verwendung immobilisierter plasmonischer Substrate, die Reproduzierbarkeit der SERS-Messung von Pollenextrakten zu erhöhen^[19] und damit eine erfolgreiche Klassifizierung von Pollenproben zu erreichen^[21]. Um die SERS-Untersuchung für die Klassifizierung von Pollenextrakten zu etablieren, ist es allerdings notwendig, die Analyse und Vorbereitung der Messung möglichst einfach zu halten. Deshalb ist es erstrebenswert, eine Untersuchung in Nanopartikellösung ohne zusätzliche Herstellung von spezifischen Substraten zu erreichen. Die durch die Verwendung einer Nanopartikellösung als Substrat gegebene, niedrigere Reproduzierbarkeit muss durch das Anpassen experimenteller Parameter ausgeglichen werden. Eine naheliegende Möglichkeit, um relativ einfach die Reproduzierbarkeit der Analyse zu erhöhen, ist eine Erhöhung der gemessenen Spektren pro Probe. Das Verändern anderer Parameter, wie z.B. der Eigenschaften der Nanopartikellösung und Nanoaggregate oder die Variation der Extraktionszeit (Ergebnisse hier nicht gezeigt) führte nicht zu reproduzierbaren Pollenspektren. Es wurden jeweils zwei Pollenextrakte von acht verschiedenen, kommerziell erhältlichen Pollenproben hergestellt und anschließend von jedem Extrakt zweimal 500 Spektren aufgenommen, die im Laufe der Spektrenvorbereitung gemittelt wurden. In einer PCA soll nun überprüft werden, ob die reproduzierbare Messung einer Pollenprobe möglich ist. In den Scores der ersten beiden Hauptkomponenten (Abbildung 4.4a) ist ersichtlich, dass die Pollenproben von *Secale cereale*, *Juniperus scopulorum*, *Artimisia absinthium* und *Sorghum halepense* jeweils große Ähnlichkeiten zueinander aufweisen und damit probenspezifische Fingerabdrücke in den Spektren vorhanden sein müssen. Die Spektren der Pollenproben von *Betula fontinalis* und *Artimisia tridentata* sind allerdings durch ähnliche Scores für die ersten beiden Hauptkomponenten charakterisiert und können deshalb anhand dieser Auftragung nicht unterschieden werden. Die Scores der dritten Hauptkomponente (Abbildung 4.4b) erlauben jedoch eine Differenzierung der Spektren dieser Proben. Die Spektren der Proben der Gattung *Populus* weisen in allen Hauptkomponenten ähnliche Scores auf und können nicht unterschieden werden. Dies ist vermutlich auf eine hohe taxonomisch bedingte Ähnlichkeit der Proben zurückzuführen. Aus den Loadings und den zugehörigen Varianzen der ersten beiden Hauptkomponenten (Abbil-

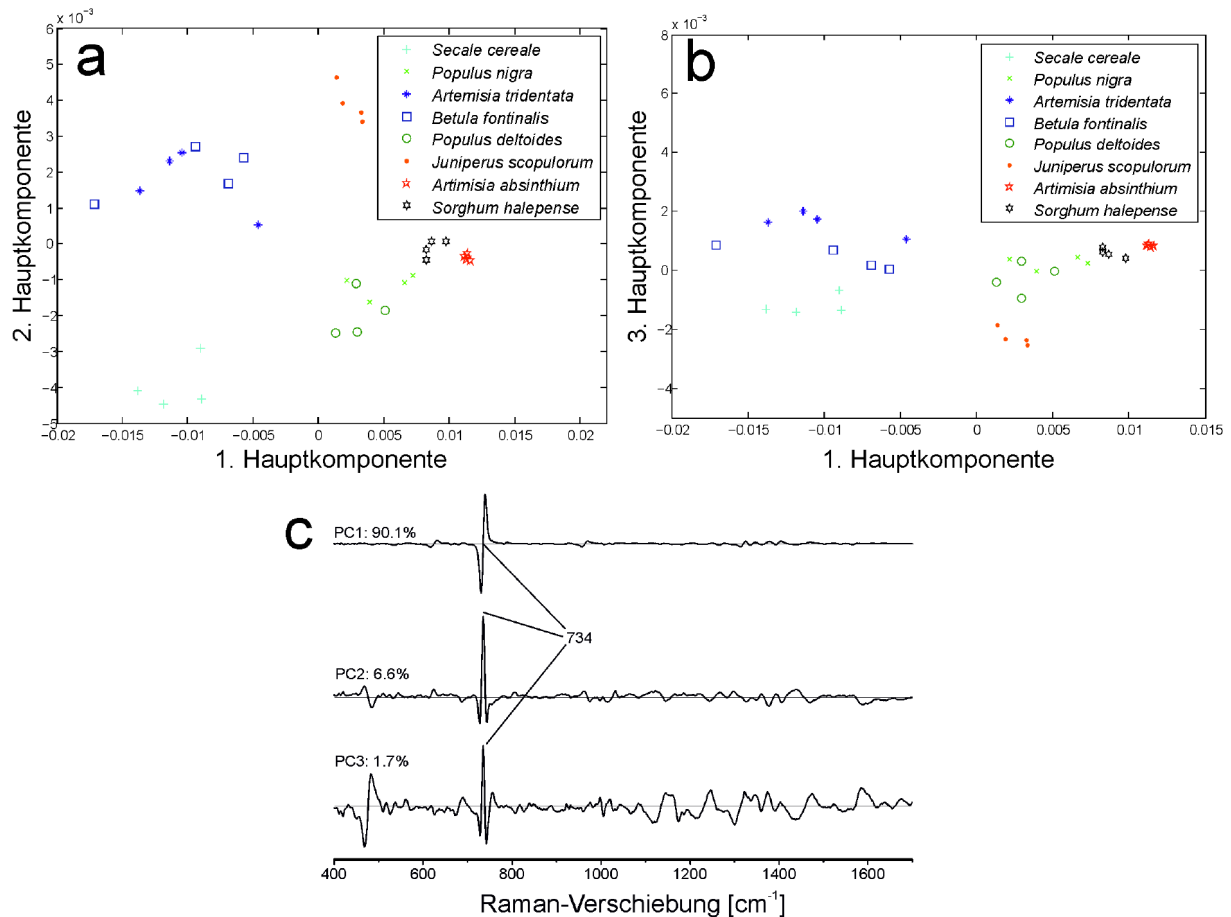


Abbildung 4.4: Ergebnis der PCA von SERS-Spektren wässriger Extrakte von kommerziell erhältlichen Pollenkörnern. Es sind die Scores der ersten und zweiten Hauptkomponente (a), die Scores der ersten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der entsprechenden Hauptkomponenten (c) gezeigt.

Abbildung 4.4c, erste und zweite Spur) ist ersichtlich, dass der größte Teil der Varianz von über 96% im Datensatz größtenteils auf Variationen der Ringatmungsschwingung bei 734 cm⁻¹ zurückzuführen ist (siehe auch Diskussion in 4.1.1). In der dritten Hauptkomponente werden jedoch auch verstärkt andere Spektralbereiche mit einbezogen (Abbildung 4.4c, dritte Spur). Dass die Pollenproben von *Betula fontinalis* und *Artemisia tridentata* nur durch diese Hauptkomponente unterschieden werden können, zeigt, dass die alleinige Auswertung der Bande der Ringatmungsschwingung nicht zielführend für die Klassifizierung verschiedener Pollenproben ist. Des Weiteren wird durch eine derartig intensive Gewichtung dieser einen Bande das Ziel, Fingerabdrücke von Pollenextrakten zu klassifizieren, nicht erreicht. Deshalb sollte in der Datenauswertung eine Möglichkeit geschaffen werden, um den Einfluss dieser Bande zu verringern und den der restlichen Spektralbereiche zu erhöhen. Um dies zu erreichen, wurden die Daten zu einer Standardnormalverteilung transformiert, was bedeutet, dass jeder Datenpunkt eines Spektrums durch die Standardabweichung dieses Datenpunktes im ganzen Datensatz geteilt wird. Die Ergebnisse der PCA von Daten unter Durchführung

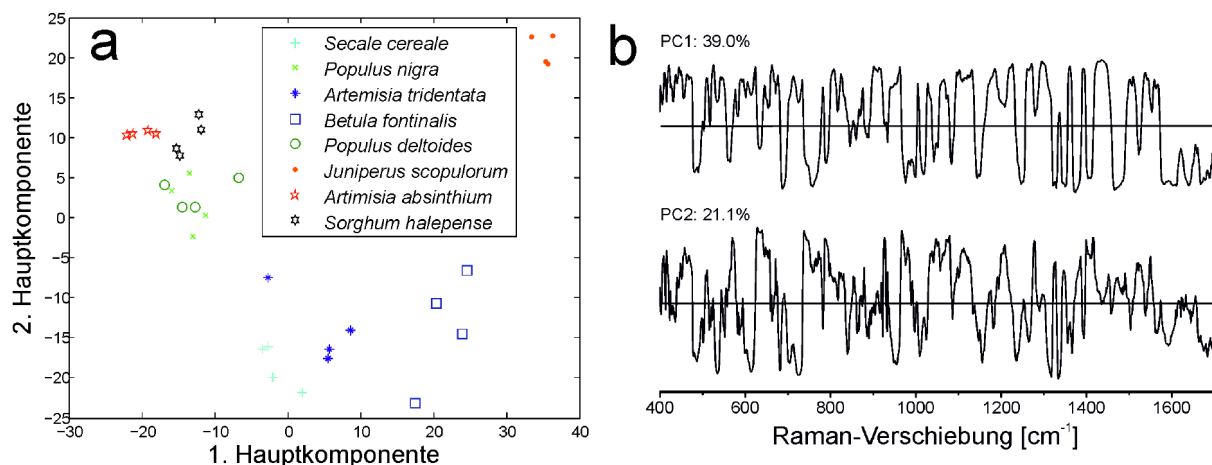


Abbildung 4.5: Ergebnis der PCA nach Transformation der Daten zu einer Standardnormalverteilung. Es sind die Scores (a), sowie die Loadings und Varianzen (b) der ersten beiden Hauptkomponenten gezeigt.

dieser zusätzlichen Normierung (Abbildung 4.5) zeigen, dass zusätzlich zu den bereits zuvor korrekt klassifizierten Spektren, die Spektren von *Betula fontinalis* und *Artemisia tridentata*, bereits durch die Auftragung der ersten beiden Hauptkomponenten (Abbildung 4.5a) voneinander unterschieden werden können. Die Loadings dieser beiden Hauptkomponenten (Abbildung 4.5b) zeigen, dass aus der Normierung eine gleichmäßige Gewichtung des gesamten Spektralbereichs resultiert. Dies führt dazu, dass die ersten beiden Hauptkomponenten, die 39.0%, bzw. 21.1% der Varianz im Datensatz repräsentieren, für die Klassifizierung der Pollenproben verwendet werden können. Die beiden Pollenproben der Gattung *Populus* können jedoch auch durch diese zusätzliche Normierung nicht voneinander unterschieden werden. Dies ist, wie zuvor bereits vermutet, wahrscheinlich in der taxonomischen Ähnlichkeit der beiden Proben begründet.

Zusammenfassend lässt sich also sagen, dass durch die Verwendung einer hinreichend großen Anzahl an Spektren pro Pollenprobe auch durch die Verwendung von Nanopartikellösung als SERS-Substrat spezifische Fingerabdrücke für die Klassifizierung erhalten werden können.

4.3 Klassifizierung von SERS-Daten mit multivariater Verfahren

Um ein geeignetes multivariates Verfahren für die Auswertung der SERS-Daten zu finden, wurde eine größere Anzahl an kommerziell erhältlichen Pollenproben untersucht. Für eine mögliche Anwendung dieser Methode ist es hierbei erstrebenswert eine möglichst geringe Menge der jeweiligen Pollenproben für die Klassifizierung verwenden zu können. Deshalb wurde zuvor die Menge an verwendeter Pollenprobe optimiert (hier nicht gezeigt). Als optimale Probenmenge für die Untersuchung wurde eine Extraktkonzentration bestimmt, die

um ca. 10% geringer war, als für die generelle Untersuchung der Eignung von SERS zur taxonomischen Klassifizierung (siehe Abschnitt 4.2). Für Pollenkörner der Art *Secale cereale* entspricht diese Konzentration etwa dem Extrakt von 100 Pollenkörnern.^[21] Die Verwendung geringerer Mengen führte dazu, dass entweder eine Klassifizierung nur durch Verwendung einer sehr hohen Anzahl an Spektren pro Probe erfolgen konnte oder gänzlich scheiterte. In den hier diskutierten Experimenten wurden jeweils viermal 500 Spektren von zwei separat hergestellten Pollenextrakten von 14 verschiedenen Pollenarten mit unterschiedlichsten taxonomischen Verwandtschaftsverhältnissen (siehe Tabelle 3.1) aufgenommen. Abbildung 4.6 zeigt die Mittelwertspektren der vektornormierten SERS-Spektren. Durch den Vergleich

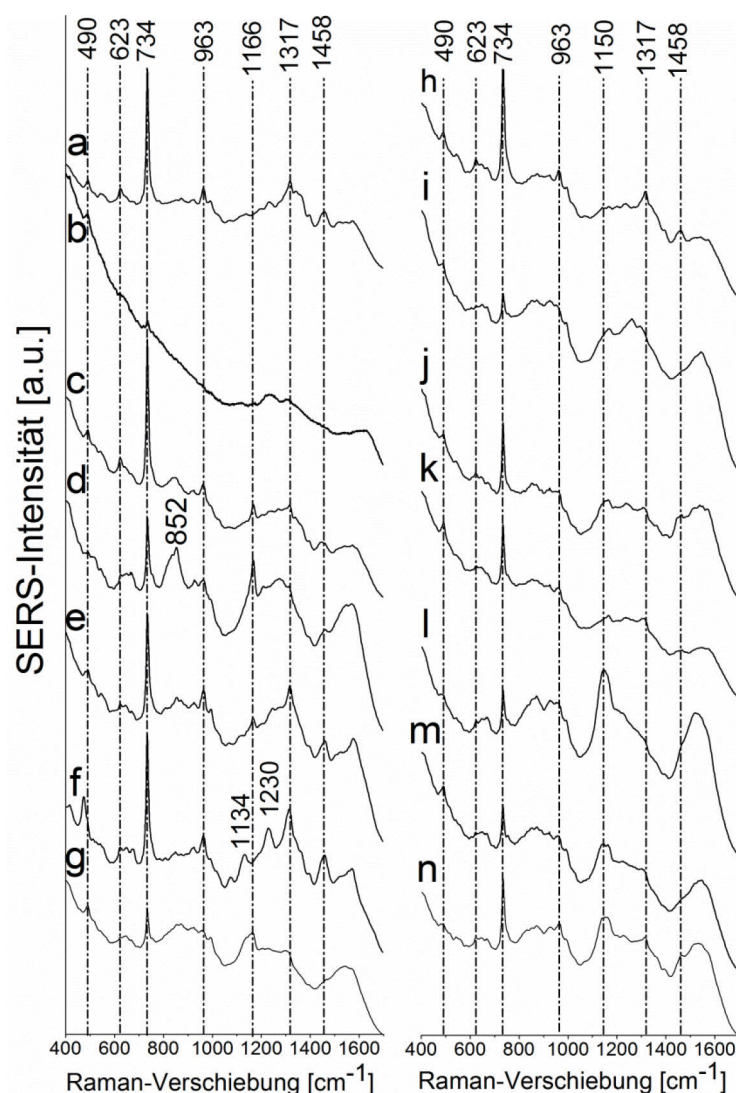


Abbildung 4.6: Mittelwertspektren von 500 vektornormierten SERS-Spektren des wässrigen Extraktes von: a; *Bassia scoparia*, b; *Artemisia absinthium*, c; *Artemisia tridentata*, d; *Iva xantifolia*, e; *Betula fontinalis*, f; *Juniperus scopulorum*, g; *Juglans nigra*, h; *Secale cereale*, i; *Sorghum halepense*, j; *Populus deltoides*, k; *Populus nigra*, l; *Populus tremuloides*, m; *Populus trichorcarpa*, n; *Ulmus pumila*.

Tabelle 4.3: Raman-Verschiebungen charakteristischer Banden in den SERS-Mittelwertspektren und deren mögliche Zuordnung zu Referenzspektren

Raman-Verschiebung [cm^{-1}]	mögliche Zuordnung ^a
1458	G,A: δ, ν^b
1363	Trp: Ring ^b
1315	C: ν^b
1230	Trp: Ringatmung ^b
1205	His: δ Ring ^c
1194	C: ν^b
1166	His: δ^c
1134	Aminosäuren NH_3^+ : δ^c
1014	Phe: Ringatmung ^b
994	Trp: asym Indolatmung ^b
963	A,G: δ^b
922	T: δ^b
890	Indol ^d
854	Ser: ν^c
827	Val: ν^c
800	C: Ringatmung ^b
758	Trp:sym Indolatmung ^b
734	A: Ringatmung ^b
662	G: Ringatmung ^b
623	A: δ^b
542	Indol ^d
490	C,G,T: δ^b

^a Abkürzungen: ν , Streckschwingung δ , Deformationsschwingung;
A, Adenin; C, Cytosin; T, Thymin; G, Guanin; His, Histidin;
Phe, Phenylalanin; Trp, Tryptophan; Val, Valin

^b nach [21]

^c nach [96]

^d nach [97]

mit Referenzspektren wurde eine mögliche Zuordnung der intensivsten Banden durchgeführt (Tabelle 4.3). Dabei kann festgestellt werden, dass wie bei den einzelnen SERS-Spektren (vgl. Abbildung 4.1 und Tabelle 4.1) ein Großteil der in den Spektren auftretenden Banden Nukleobasen und Aminosäuren zugeordnet werden können. Die zuvor thematisierte große Variation der Intensität der Ringatmungsschwingung bei 734 cm^{-1} zwischen verschiedenen Arten (vgl. Abschnitt 4.2) ist auch in diesen Spektren erkennbar.

Beim Vergleich der Spektren in Abbildung 4.6 können artspezifische Banden in den Spektren identifiziert werden. Beispielsweise liefern *Juniperus scopulorum*-Pollenextrakte Banden bei 1134 und 1230 cm^{-1} , die einer NH_3^+ -Deformationsschwingung von Aminosäuren, bzw. einer

Ringatmungsschwingung von Tryptophan zugeordnet werden können (siehe Tabelle 4.3). Diese Banden sind in den Spektren von Pollenextrakten anderer Proben entweder sehr viel geringer ausgeprägt oder treten gar nicht auf.

Um ein geeignetes Verfahren zum Klassifizieren der Pollenextraktspektren zu finden, sollen im Folgenden sowohl die unüberwachten, häufig verwendeten Methoden der hierarchischen Clusteranalyse (HCA) und Hauptkomponentenanalyse (*Principal Component Analysis*, PCA), als auch die für SERS-Daten bisher nur vereinzelt eingesetzte, überwachte Methode der künstlichen neuronalen Netze (*artificial neural networks*, ANN) erprobt werden. Um Spektren auszuwerten, die eine möglichst große Anzahl an Bestandteilen des Extraktes repräsentieren, wurden für die Auswertungen mit unüberwachten Methoden erste Ableitungen von Mittelwertspektren der jeweils 8 Präparationen pro Pollenprobe verwendet. Eine Übergewichtung der Ringatmungsschwingung wurde verhindert, indem jeweils eine Transformierung der einzelnen Datenpunkte zu einer Standardnormalverteilung durchgeführt wurde (vgl. Abschnitt 4.2). Da für jede Art jeweils zwei Extrakte untersucht wurden, kann bei der Auswertung auch analysiert werden, ob sich die verschiedenen Extrakte unterscheiden und damit, inwiefern ggf. vorhandene Unterschiede in der Probenpräparation das Klassifizierungsergebnis beeinflussen.

4.3.1 HCA von SERS-Spektren wässriger Pollenextrakte

In einer hierarchischen Clusteranalyse (HCA) soll überprüft werden, ob euklidische Distanzen zwischen verschiedenen Pollenextraktspektren für die Klassifizierung verwendet werden können. Das Ergebnis der HCA (Abbildung 4.7) zeigt, dass Spektren der Pollenextrakte *Bassia scoparia*, *Secale cereale*, *Sorghum halepense*, *Artemisia absinthium* und *Ulmus pumila* Cluster bilden, die jeweils ausschließlich Spektren der jeweiligen Art enthalten. Dies deutet darauf hin, dass die Spektren dieser Arten eine sehr geringe Varianz aufweisen, obwohl diese durch Untersuchung verschiedener Extrakte erhalten wurden. Die Spektren von *Populus deltoides*, *Populus nigra*, *Populus tremuloides* und *Populus trichocarpa* bilden drei Cluster, die Spektren verschiedener Arten und Extrakte enthalten (*Populus spec.* in Abbildung 4.7). Die verschiedenen Arten der Gattung *Populus* können demnach durch eine HCA nicht unterschieden werden. Die Spektren von *Juniperus scopulorum*, *Betula fontinalis*, *Artemisia tridentata*, *Iva xantifolia* und *Juglans nigra* weisen jeweils zwei unterschiedliche spektrale Klassen auf, die in jeweils zwei unterschiedlichen Clustern der jeweiligen Arten resultieren. Die unterschiedlichen Cluster der jeweiligen Art (in der Abbildung gekennzeichnet durch die Nummerierungen: (1) und (2)) können hauptsächlich der Untersuchung jeweils eines

Extraktes zugeordnet werden. Dies deutet darauf hin, dass im Datensatz große Varianzen vorhanden sind, die auf unterschiedliche, chemische Zusammensetzungen einzelner Extrakte zurückgeführt werden können. Um eine art-, bzw. probenspezifische Auswertung zu erreichen, ist es also notwendig, eine Trennung zwischen präparationsbedingten und art-, bzw. probenbedingten Varianzen zu ermöglichen.

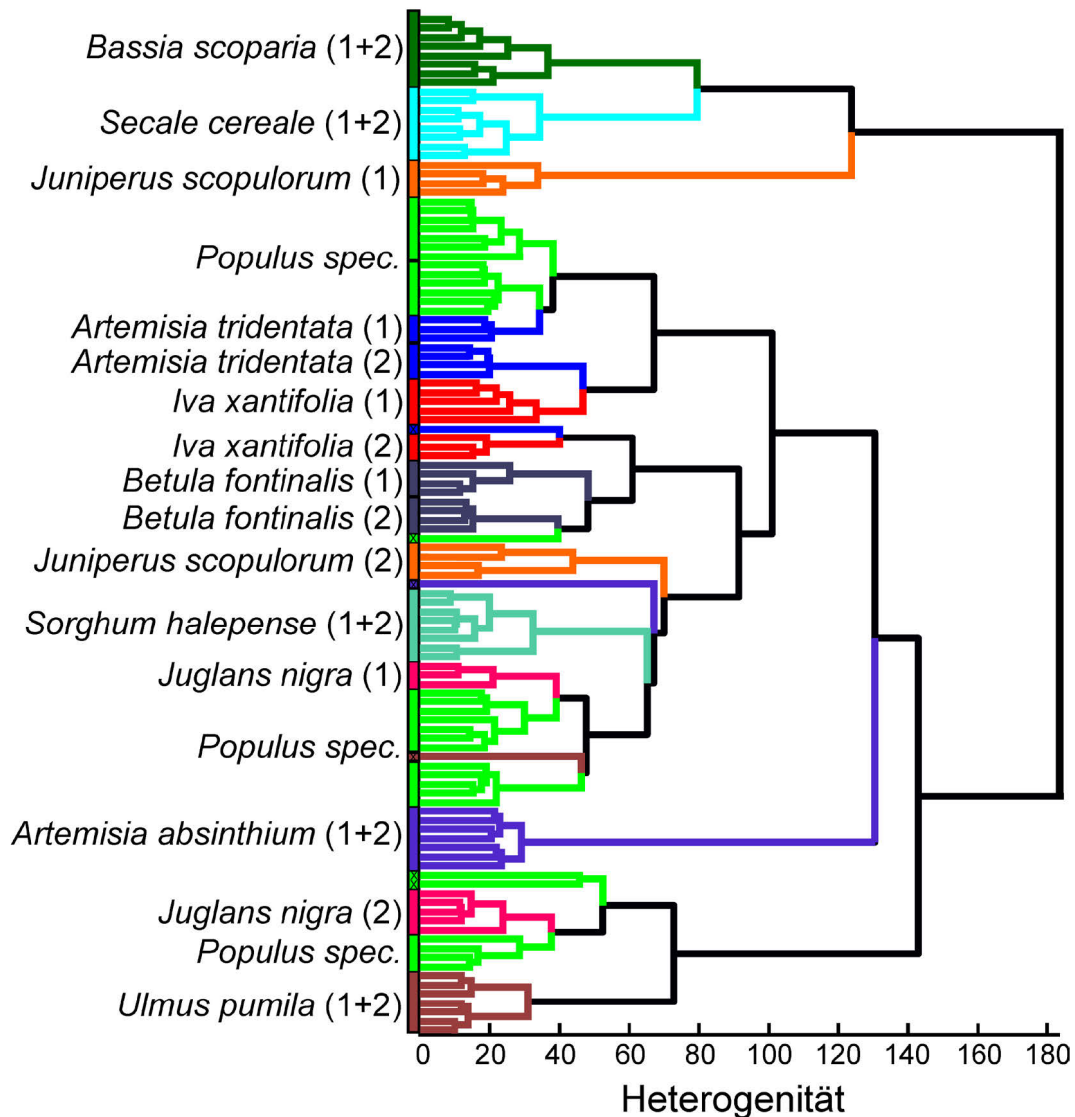


Abbildung 4.7: Ergebnis der hierarchischen Clusteranalyse von 112 Mittelwertspektren (jeweils acht Mittelwertspektren von 14 verschiedenen Pollenarten) im Spektralbereich zwischen 400 und 1700 cm^{-1} . Es wurden jeweils 500 vektornormierte, abgeleitete Spektren gemittelt und die spektralen Daten zu einer Standardnormalverteilung transformiert. Zur besseren Übersichtlichkeit sind die Spektren von jeweils einer Art in gleichen Farben, aber anderer Nummerierung dargestellt. Die Nummerierung kennzeichnet verschiedene Cluster derselben Art, die häufig den einzelnen Extrakten, die jeweils für jede Art untersucht wurden, entsprechen.

4.3.2 PCA von SERS-Spektren wässriger Pollenextrakte

Eine Möglichkeit, die verschiedenen Varianzen im Datensatz zu separieren, könnte die Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) bieten, weshalb die SERS-Daten anschließend mit dieser multivariaten Methode ausgewertet wurden. Die Scores, Loadings und Varianzen der ersten vier Hauptkomponenten, die 64% der Gesamtvarianz im Datensatz repräsentieren, sind in Abbildung 4.8 gezeigt. Die Loadings (Abbildung 4.8a) zeigen, dass durch die Transformation der Daten zu Standardnormalverteilung eine gleichmäßige Gewichtung des ganzen Spektralbereiches erreicht wird. Die Kombination der Scores der ersten vier Hauptkomponenten (Abbildung 4.8b bis d) ermöglicht eine Unterscheidung der Pollenproben, die auch mit HCA unterschieden werden können (vgl. Abschnitt

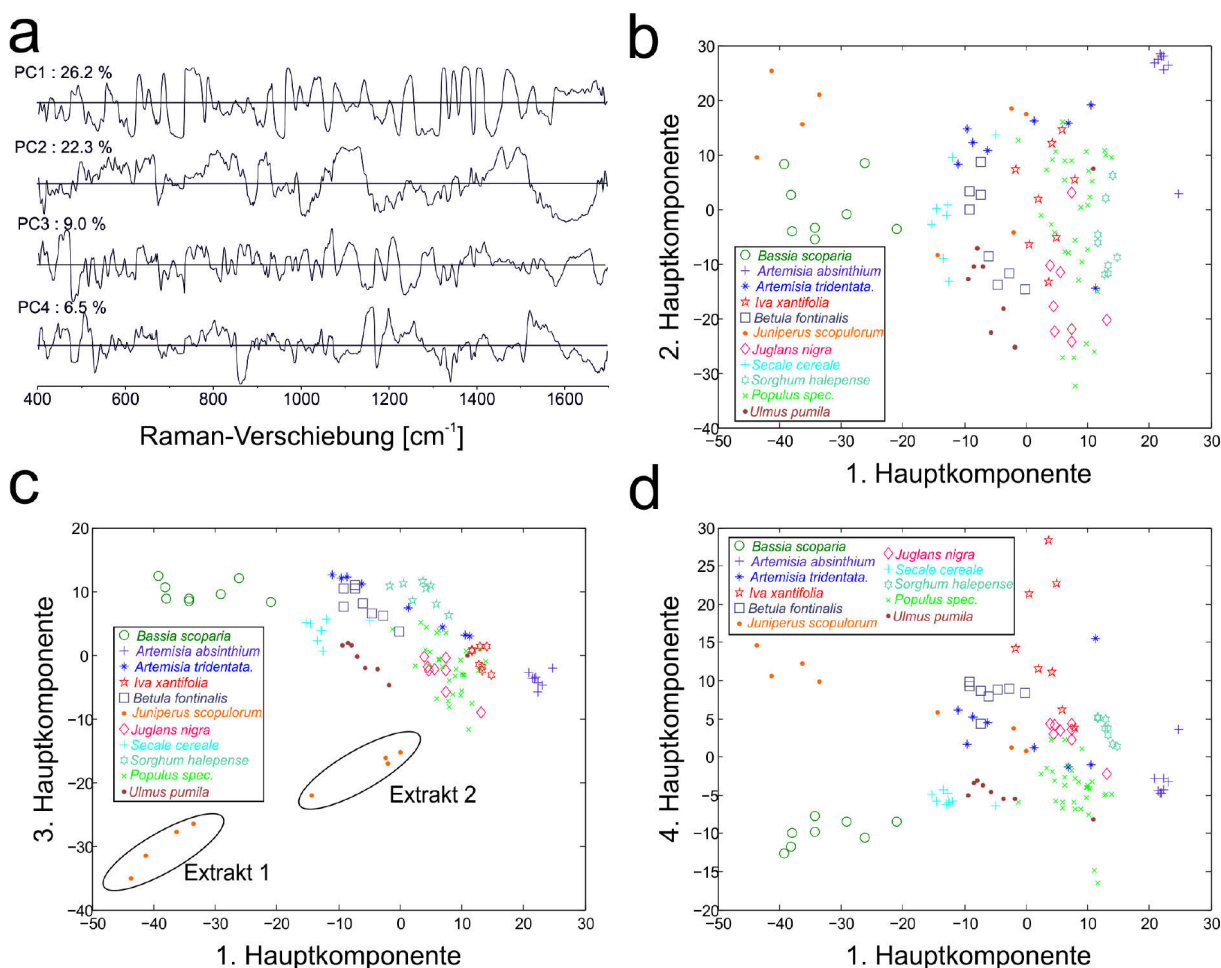


Abbildung 4.8: Ergebnis der PCA von 112 Mittelwertspektren (jeweils acht Mittelwertspektren von 14 verschiedenen Pollenarten) im Spektralbereich zwischen 400 und 1700 cm^{-1} . Loadings der ersten vier Hauptkomponenten (a), sowie die Scores der ersten und zweiten (b), ersten und dritten (c) und ersten und vierten Hauptkomponente (d) sind gezeigt. Es wurden jeweils 500 vektornormierte, abgeleitete Spektren gemittelt und die spektralen Daten zu einer Standardnormalverteilung transformiert.

4.3.1). Zusätzlich können durch die Scores der ersten, dritten und vierten Hauptkomponente alle Spektren der Arten *Juniperus scopulorum*, *Betula fontinalis* und *Iva xantifolia* von den Spektren der anderen Arten separiert werden (Abbildung 4.8c und d). Für die Spektren von *Artemisia tridentata*, *Juglans nigra* und den verschiedenen *Populus*-Arten ist dies allerdings auch durch Auswertung der Daten mit PCA nicht möglich (Abbildung 4.8b bis d). Betrachtet man die Scores der ersten Hauptkomponente, so ist ersichtlich, dass diese für die Differenzierung zwischen Spektren verschiedener Pollenarten entscheidend sind. Dies lässt sich z.B. anhand der Spektren von *Sorghum halepense* oder *Artemisia absinthium* erkennen (Abbildung 4.8b bis d). Die entsprechenden Scores der Pollenspektren von *Juniperus scopulorum* (Abbildung 4.8c, orangefarbene Punkte) weisen allerdings sehr unterschiedliche Werte auf, die dazu führen, dass nur anhand dieser Hauptkomponente keine eindeutige Klassifizierung durchgeführt werden kann. Bezieht man die Information hinsichtlich der Zuordnung der Spektren zu Extrakten in die Klassifizierung mit ein (Abbildung 4.8c), können die spektralen Varianzen dieser Art hauptsächlich auf solche zurückgeführt werden, die zwischen verschiedenen Extrakten auftreten. Dies verdeutlicht, dass auch unter Anwendung einer PCA die präparationsbedingten nicht von den probenbedingten Varianzen getrennt werden können.

4.3.3 Unterscheidung verschiedener Pollenproben durch Auswertung mit ANN

Da es durch die bisher angewendeten, unüberwachten Verfahren nicht möglich ist, die Varianzen von Spektren verschiedener Arten aus den SERS-Daten herauszuarbeiten, sollen nun künstliche neuronale Netz (*artificial neural networks*, ANN) für die Auswertung erprobt werden. Diese sind im Vergleich zu anderen überwachten Mustererkennungsverfahren, wie zum Beispiel support vector machines (SVM), die ebenfalls erfolgreich für die Klassifizierung von SERS-Spektren eingesetzt wurden,^{[94],[95]} besonders leistungstark, wenn eine Klassifizierung in mehr als zwei Klassen erfolgen soll.^[81] Es wurden jeweils 894 spektrale Datenpunkte von 2000 Spektren der vier Untersuchungen eines Extraktes als Eingabeschicht für das Trainieren und Validieren des Netzes verwendet, während die 2000 Spektren des jeweils anderen Extraktes ausnahmslos für das Testen benutzt wurden. Die Ergebnisse (Abbildung 4.9) zeigen, dass eine korrekte Identifizierung bei 97% der Proben möglich ist. Lediglich ein Datensatz von *Populus deltoides* und einer von *Populus nigra* werden der falschen Klasse zugeordnet. Auch die Spektren verschiedener Arten der Gattung *Populus* können mit ANN unterschieden werden. Da jeweils die Spektren unterschiedlicher Extrakte für das Trainieren und Testen des ANN verwendet wurden, muss die Klassifizierung

identifiziert als	Bassia scoparia	Artemisia absinthium	Artemisia tridentata	Iva xanthifolia	Betula fontinalis	Juniperus scopulorum	Juglans nigra	Secale cereale	Sorghum halepense	Populus deltoides	Populus nigra	Populus tremuloides	Populus trichocarpa	Ulmus pumila
Art der getesteten Spektren														
Bassia scoparia	4(100%)													
Artemisia absinthium		4(100%)												
Artemisia tridentata			4(100%)											
Iva xanthifolia				4(100%)										
Betula fontinalis					4(100%)									
Juniperus scopulorum						4(100%)								
Juglans nigra							4(100%)							
Secale cereale								4(100%)						
Sorghum halepense									4(100%)					
Populus deltoides			1(25%)							3(75%)				
Populus nigra				1(25%)							3(75%)			
Populus tremuloides												4(100%)		
Populus trichocarpa													4(100%)	
Ulmus pumila														4(100%)

Abbildung 4.9: Ergebnis der ANN-Klassifizierung von SERS-Daten von 14 verschiedenen, kommerziell erhältlichen Pollenarten. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

der SERS-Daten auf den probenspezifischen Beiträgen der SERS-Spektren basieren und die präparationsbedingten Varianzen konnten durch diese Auswertung erfolgreich vernachlässigt werden. Damit wurde gezeigt, dass auch SERS-Daten zur Analyse spezifischer, spektraler Fingerabdrücke von komplexen biologischen Proben mit ANN verwendet werden können. Für die Auswertung anderer schwingungsspektroskopischer Methoden wurde dies bereits z.B. anhand von der Klassifizierung von Bakterien^[35], der Identifizierung verschiedener Malaria Zellzyklen^[103] oder der Entwicklung bildgebender Verfahren für Gewebeuntersuchungen^[104] gezeigt.

4.4 Wiedergabe taxonomischer Verwandtschaftsverhältnisse aus SERS-Daten

In den vorigen Abschnitten wurde anhand von kommerziell erhältlichen Pollenproben gezeigt, dass mit SERS des wässrigen Pollenextraktes Fingerabdrücke der inneren Pollenbestandteile erhalten werden können. Diese können durch Kombination mit neuronalen Netzen für eine spezifische Klassifizierung der Pollenproben verwendet werden. Um zu untersuchen, ob sich taxonomische Verwandtschaftsverhältnisse aus den SERS-Daten reproduzieren lassen, wurden ausgewählte Pollenproben in Berliner Parks und im Botanischen Garten Berlin-Dahlem gesammelt und mit SERS analysiert. Es wurden Pollenproben von 70 verschiedenen Arten und zehn Gattungen der Ordnungen *Fagales* und *Coniferales* (siehe Tabelle 3.2) untersucht. Dabei wurden auf gleiche Weise wie in Abschnitt 4.3 jeweils von jeder Probe zweimal 500 SERS-Spektren aufgenommen und für die anschließende Klassifizierung verwendet.

identifiziert als Ordnung der getesteten Spektren	Coniferales	Fagales
Coniferales	104(95%)	6(5%)
Fagales	9(11%)	73(89%)

Abbildung 4.10: Ergebnis der ANN-Klassifizierung von 110 Datensätzen der Ordnung *Coniferales* und 82 Datensätzen der Ordnung *Fagales*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

Abbildung 4.10 zeigt das Ergebnis der Klassifizierung auf der taxonomischen Ebene der Ordnung. Es werden ca. 92% der Spektren der jeweils korrekten Ordnung zugewiesen. Daraus kann geschlossen werden, dass eine Klassifizierung der taxonomischen Ordnung mit ANN der SERS-Daten in den meisten Fällen erfolgreich durchgeführt werden kann. Dies zeigt, dass die taxonomische Klassifizierung durch SERS auch bei heterogenen taxonomischen Gruppen erfolgreich ist. Um SERS-Daten für die Klassifizierung von Gattungen zu verwenden, wurden anschließend die Spektren der Ordnungen *Fagales* und *Coniferales* jeweils separat mit ANN untersucht. Die Ergebnisse der Klassifizierungen sind in den Abbildung 4.11 und Abbildung 4.12 gezeigt. Insgesamt werden ca. 78% der Proben der Ordnung *Coniferales* und ca. 80% der Proben der Ordnung *Fagales* korrekt zugeordnet. Für die Gattungen *Taxus*, *Picea*, *Pinus*, *Alnus*, *Betula*, *Quercus* und *Corylus* werden dabei jeweils mindestens 75% der Datensätze korrekt klassifiziert. Die meisten Datensätze der Gattungen *Abies*, *Juniperus* und *Myrica* werden jedoch jeweils anderen Gattungen zugewiesen. Da bei diesen Gattungen jeweils nur 6 Pollenproben und damit auch nur jeweils 6 Datensätze zum Trainieren und 6 Datensätze zum

identifiziert als Gattung der getesteten Spektren	Taxus	Picea	Juniperus	Abies	Pinus
Taxus	13 (81%)	2 (13%)			1 (6%)
Picea		20 (77%)			6 (23%)
Juniperus	1 (17%)	3 (50%)	2 (33%)		
Abies		3 (50%)		0 (0%)	3 (50%)
Pinus		5 (9%)			51 (91%)

Abbildung 4.11: Ergebnis der ANN-Klassifizierung von 16 Datensätzen der Gattung *Taxus*, 26 Datensätzen der Gattung *Picea*, jeweils sechs Datensätzen der Gattungen *Juniperus* und *Abies*, sowie 56 Datensätzen der Gattung *Pinus*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

identifiziert als Gattung der getesteten Spektren	<i>Alnus</i>	<i>Betula</i>	<i>Quercus</i>	<i>Myrica</i>	<i>Corylus</i>
<i>Alnus</i>	18 (100%)				
<i>Betula</i>		15 (75%)	1 (5%)		4 (20%)
<i>Quercus</i>		1 (5%)	17 (85%)		2 (10%)
<i>Myrica</i>		3 (50%)	2 (33%)	0 (0%)	1 (17%)
<i>Corylus</i>		2 (11%)			16 (89%)

Abbildung 4.12: Ergebnis der ANN-Klassifizierung von 18 Datensätzen der Gattung *Alnus*, 20 Datensätzen der Gattung *Betula*, 20 Datensätzen der Gattung *Quercus*, sechs Datensätze der Gattung *Myrica*, sowie 18 Datensätzen der Gattung *Corylus*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

Testen des ANN vorhanden waren, ist davon auszugehen, dass der Grund für die mangelnde Klassifizierung vor allem in der geringen Anzahl der Daten zu suchen ist. Da bei der Untersuchung der kommerziell erhältlichen Pollenproben (siehe Abschnitt 4.3.3) in den meisten Fällen die Spektren eines Extraktes ausreichend waren, um die Spektren eines anderen Extraktes derselben Probe korrekt zu klassifizieren, ist es wahrscheinlich, dass die Ursache der nicht korrekten Klassifizierung hier in der größeren Diversität der Proben der jeweiligen Gattungen liegt. Insgesamt wurden dennoch ca. 80% der Proben der korrekten Gattung zugeordnet. Dies deutet darauf hin, dass bei einer ausreichenden Anzahl an Datensätzen auch eine korrekte Klassifizierung der Gattung aus den SERS-Daten erfolgen kann.

Um zu überprüfen, ob sich taxonomische Verwandtschaftsverhältnisse auch auf der taxonomischen Ebene der Arten aus den SERS-Daten reproduzieren lassen, wurden Spektren von *Pinus mugo* und *Pinus sylvestris*, von denen jeweils eine sehr hohe Zahl an Pollenproben vorlag, untersucht. Abbildung 4.13 zeigt die Ergebnisse dieser Analyse. Es wurde jeweils nur ein Datensatz der jeweils falschen Art zugewiesen und damit eine korrekte Klassifizierung von 92,5% der Proben erreicht. Sind genügend Datensätze zum Training des ANN vorhanden, kann also auch eine Klassifizierung auf der taxonomischen Ebene der Art erreicht werden.

Die Untersuchungen von Pollenproben aus dem Feld zeigen, dass SERS in Kombination mit künstlichen neuronalen Netzen für die Klassifizierung verschiedener taxonomischer Gruppen angewendet werden kann. Es können dabei sowohl sehr heterogene Gruppen, wie zum Beispiel Proben unterschiedlicher Ordnungen unterschieden werden, als auch Proben, die sich sehr ähnlich sind, wie zum Beispiel verschiedene Arten einer Gattung.

Art der getesteten Spektren	identifiziert als	
	<i>Pinus mugo</i>	<i>Pinus sylvestris</i>
<i>Pinus mugo</i>	13 (93%)	1 (7%)
<i>Pinus sylvestris</i>	1 (8%)	11 (92%)

Abbildung 4.13: Ergebnis der ANN-Klassifizierung von 14 Datensätzen der Art *Pinus mugo* und zwölf Datensätzen der Art *Pinus sylvestris*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

4.5 Klassifizierung vorselektierter Spektren des Datensatzes

4.5.1 Vorselektion durch Extraktion mit dem ANN

Um SERS-Daten wässriger Pollenextrakte erfolgreich zu klassifizieren, müssen die probenspezifischen, taxonomischen Informationen im Datensatz von den präparationsbedingten Varianzen separiert werden (siehe Abschnitt 4.3). Dass dies durch Auswertung mit ANN möglich ist, konnte sowohl anhand von Daten kommerziell erhältlicher Pollenproben (siehe Abschnitt 4.3.3), als auch anhand derer von Pollenproben aus dem Feld (siehe Abschnitt 4.4) gezeigt werden. Bei der Klassifizierung wurden dabei die Zuordnungen der einzelnen Spektren zu den jeweiligen taxonomischen Gruppen nicht für jedes Spektrum einzeln bewertet, sondern gemittelt und anschließend anhand des jeweils größten Wertes eine Zuordnung des Datensatzes zu einer Gruppe durchgeführt. Um die für die Klassifizierung relevanten Spektren aus dem Datensatz zu extrahieren, muss die Zuordnungen der einzelnen Spektren jedoch einzeln betrachtet und bewertet werden. Es wurden deshalb erneut neuronale Netze trainiert. Da das Ziel hierbei nicht das Testen von unabhängigen Spektren für die Validierung des Klassifizierungserfolgs ist, konnten jeweils für die taxonomischen Gruppen *Abies*, *Juniperus* und *Myrica*, für die zuvor eine sehr geringe Anzahl verschiedener Datensätze vorlag (siehe Abschnitt 4.4), jeweils alle vorhanden Datensätze für das Trainieren des Netzes verwendet werden. Das neuronale Netz wurde anschließend dazu verwendet, die 500 Spektren jedes Datensatzes zu bewerten und zu unterscheiden zwischen Spektren, die für die Klassifizierung relevant sind und solchen, die bei einer Klassifizierung vernachlässigt werden können. Dabei wurden zwei prinzipiell unterschiedliche Extraktionen durchgeführt: (siehe Abbildung 4.14)

(i) Es wurden jeweils nur die Spektren extrahiert, die eine hohe Zuordnung zur jeweils korrekten Klasse aufzeigten (rote, bzw. blaue Balken in Abbildung 4.14). Damit erhält man

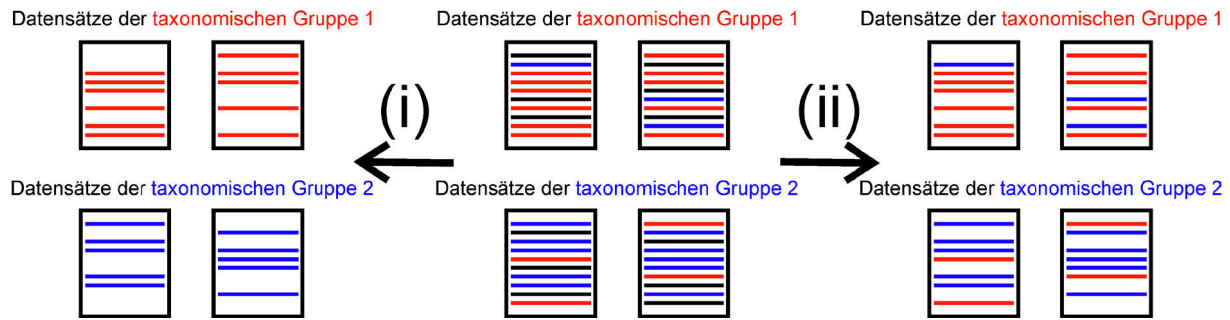


Abbildung 4.14: Schematische Darstellung der angewendeten Extraktionen.

Daten, welche die für die Klassifizierung relevanten spektralen Eigenschaften verstärkt enthalten. Weil die Grundlage für ihre Extraktion die korrekte taxonomische Einordnung durch das ANN ist, werden diese Spektren im Folgenden als *taxonomisch spezifisch* bezeichnet.

(ii) Eine zweite Extraktionsweise wurde angewendet, um diejenigen Spektren zu extrahieren, die eine hohe Zuordnung zu einer der taxonomischen Gruppen aufwies (jeweils rote und blaue Balken in Abbildung 4.14). Damit eliminiert man die Spektren, die allgemein nicht für die taxonomische Klassifizierung relevant sind (schwarze Balken in Abbildung 4.14). Die auf diese Weise extrahierten Spektren werden deshalb im Folgenden als *taxonomisch relevant* bezeichnet.

Durch die Untersuchung extrahierter Daten soll nun die in Abschnitt 4.4 begonnene Diskussion der Klassifizierung auf verschiedenen taxonomischen Ebenen weiter vertieft werden.

4.5.2 Untersuchung taxonomisch spezifischer Daten der Ordnungen

Für die Untersuchung der SERS-Daten zur Klassifizierung auf taxonomischer Ebene der Ordnung wurden zunächst alle Spektren, die eine niedrigere Zuordnung als 0.9 zur jeweiligen Klasse aufwiesen aus der anschließenden Auswertung ausgeschlossen. Daraufhin wurden die Daten eines Datensatzes gemittelt und mit HCA und PCA analysiert. Das Ergebnis der HCA (Abbildung 4.15) zeigt, dass eine klare Unterscheidung aller Datensätze zwischen den Spektren der Ordnungen *Coniferales* und *Fagales* möglich ist und dass somit eine Extraktion von Spektren, die Informationen zur Unterscheidung der taxonomischen Ordnung enthalten, erfolgreich durchgeführt werden konnte. Auch durch die erste Hauptkomponente der PCA (Abbildung 4.16) ist eine Unterscheidung zwischen den taxonomischen Ordnungen durch weitestgehend positive Scores der Spektren von *Coniferales* und negative Scores der Spektren von *Fagales* möglich. Betrachtet man die Loadings, so ist erkennbar, dass diese

Unterscheidung auf eine Vielzahl von Banden zurückgeführt werden kann. Viele können dabei DNA- oder Peptidbestandteilen zugeordnet werden (vgl. Tabelle 4.1).

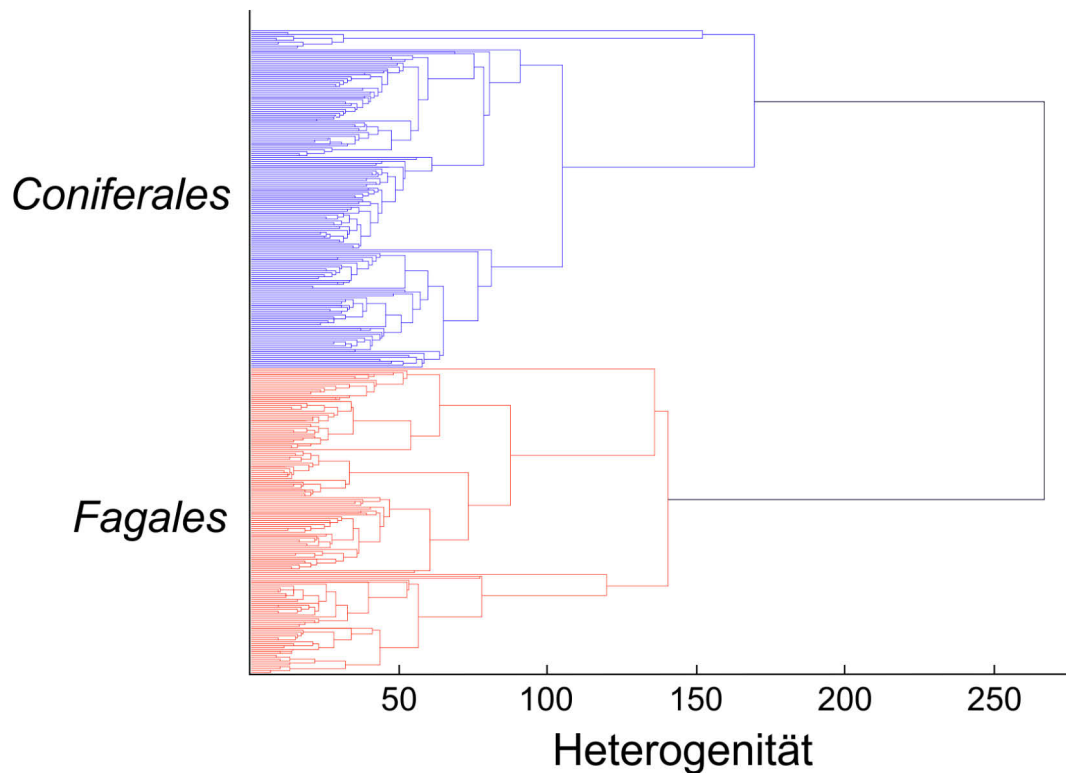


Abbildung 4.15: Ergebnis der Clusteranalyse von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden.

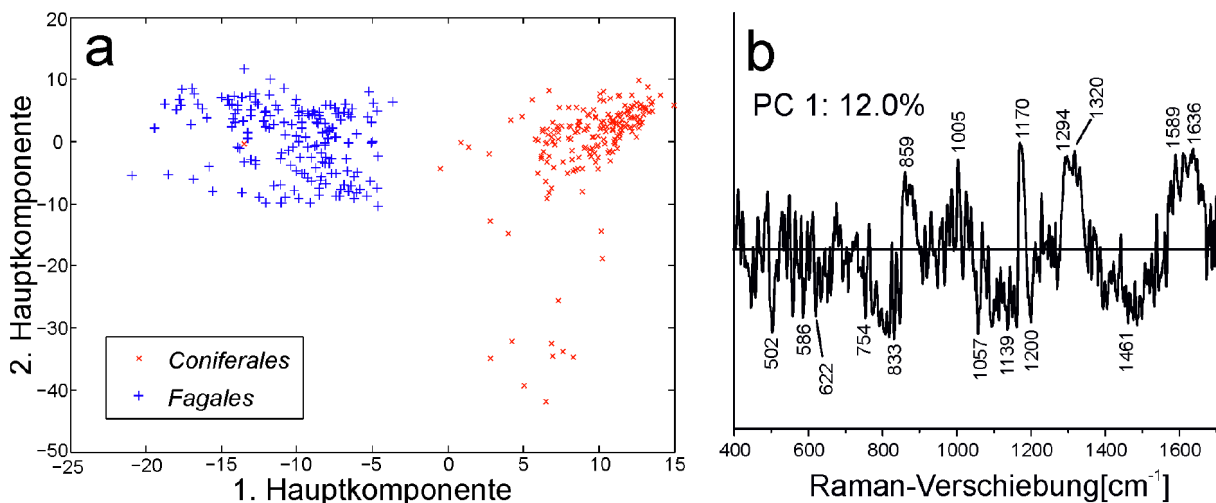


Abbildung 4.16: Ergebnis der PCA von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Die Scores der ersten beiden Hauptkomponenten (a), sowie die Loadings der ersten Hauptkomponente (b) sind gezeigt.

4.5.3 Untersuchung taxonomisch relevanter Daten der Ordnungen

Anschließend wurden jeweils alle Spektren extrahiert, die eine Zuordnung von mindestens 0.9 zu einer der beiden Klassen aufwiesen. Abbildung 4.17 zeigt das Ergebnis der HCA. Es ist erkennbar, dass eine Klassifizierung auf taxonomischer Ebene der Ordnung möglich ist, da sich ein Großteil der Mittelwertspektren (91.4%) in dem Cluster der jeweils korrekten Ordnung befindet. Dies wird auch durch die Scores der PCA bestätigt (Abbildung 4.18a), da ähnlich wie bei der Auswertung taxonomisch spezifischer Daten (vgl. Abbildung 4.16a) eine Unterscheidung in überwiegend positive Scores der Datensätze der Ordnung *Coniferales* und überwiegend negative Scores der Datensätze der Ordnung *Fagales* erfolgt. Da sich auch die Loadings der jeweiligen Hauptkomponentenanalysen (vgl. Abbildung 4.18b mit Abbildung 4.16b) ähneln, lässt sich daraus schließen, dass ein Großteil der für die jeweilige Ordnung extrahierten Spektren jeweils gleich sind und somit die Unterscheidung in verschiedene taxonomische Ordnungen anhand vollständiger SERS-Spektren möglich ist. Eine Gewichtung einzelner spektraler Bestandteile der einzelnen SERS-Spektren durch das ANN ist somit nicht zwingend für eine erfolgreiche Differenzierung zwischen *Fagales* und *Coniferales*-Pollenproben erforderlich.

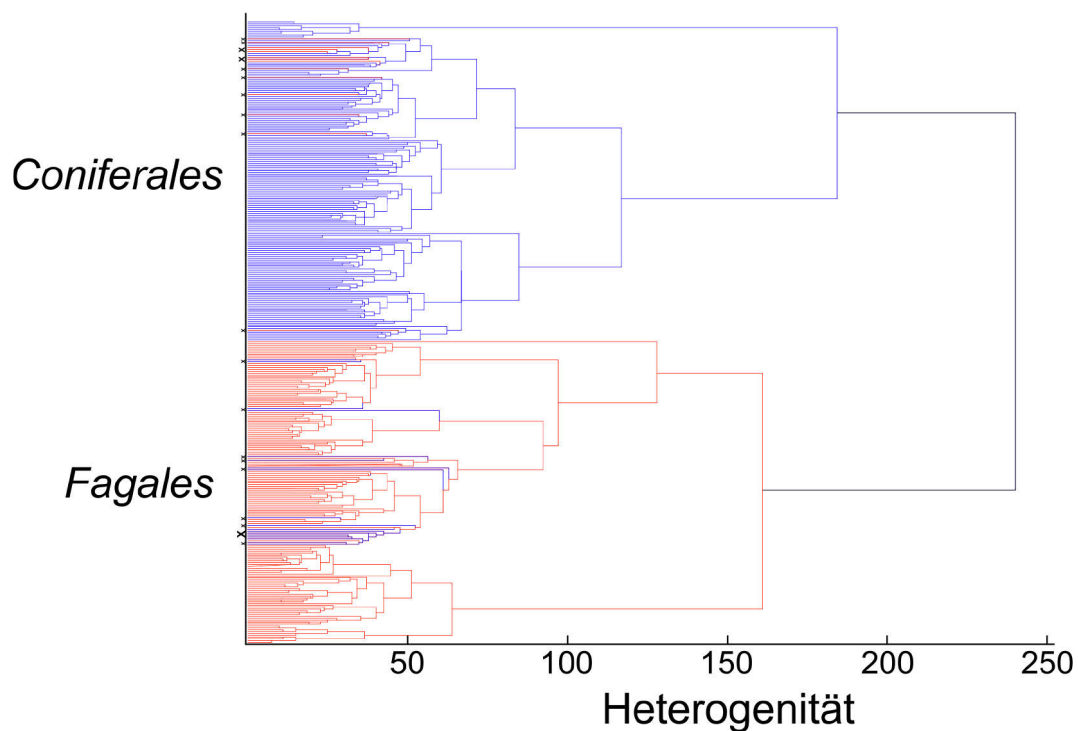


Abbildung 4.17: Ergebnis der Clusteranalyse von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden.

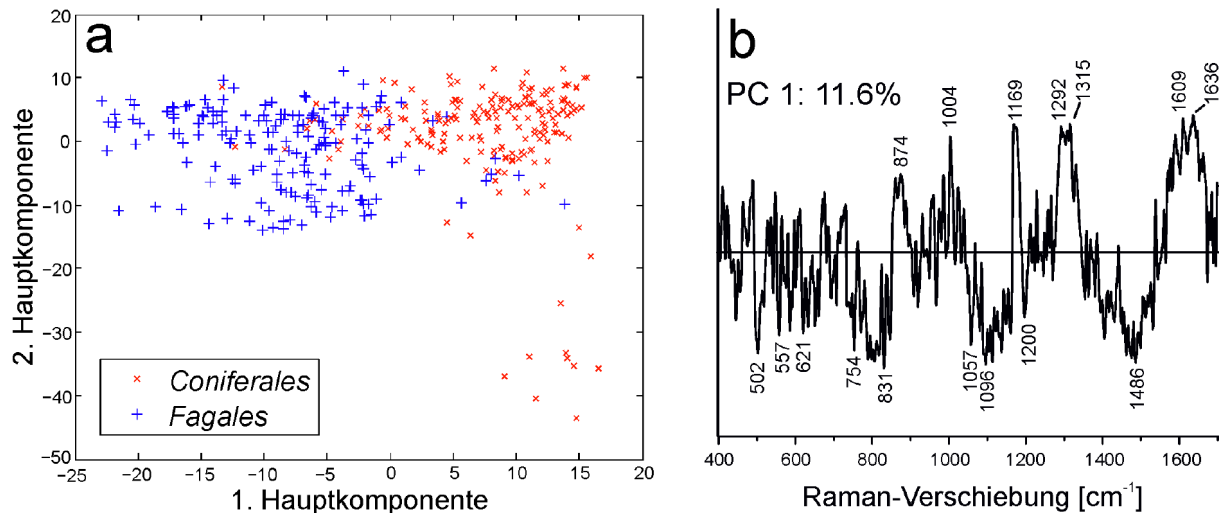


Abbildung 4.18: Ergebnis der PCA von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Die Scores der ersten beiden Hauptkomponenten (a), sowie die Loadings der ersten Hauptkomponente (b) sind gezeigt.

4.5.4 Untersuchung taxonomisch spezifischer Daten der Gattungen

Es wurden für beide Ordnungen jeweils neuronale Netze zur Extraktion der gattungsspezifischen Varianzen trainiert. Zunächst wurden dabei die Datensätze, die jeweils eine geringere Zuordnung als 0.5 zur jeweiligen Gattung aufwiesen aus der Auswertung ausgeschlossen und eine HCA der Mittelwertspektren durchgeführt. Die Ergebnisse (Abbildung 4.19) zeigen, dass gattungsspezifische Cluster gebildet werden, bei denen insgesamt 92% der Spektren dem jeweils korrekten Cluster zugeordnet werden. Es lässt sich also feststellen, dass Spektren extrahiert werden konnten, die die gattungsspezifischen Varianzen enthalten. Es sind allerdings auch Spektren vorhanden, die jeweils dem falschen Cluster zugeordnet werden. Dies betrifft einzelne Spektren verschiedener Gattungen, sowie jeweils eine Gruppierung von neun *Betula*- (Abbildung 4.19b), bzw. sechs *Picea*- und drei *Juniperus*-Spektren (Abbildung 4.19a). Allgemein weisen auch die gattungsspezifischen Cluster eine relativ hohe Heterogenität auf, was zusätzlich darauf hindeutet, dass sich weiterhin Varianzen im Datensatz befinden, die nicht für die Identifizierung der jeweiligen Gattung relevant sind. Allgemein können die Ergebnisse der HCA durch die Ergebnisse der Hauptkomponentenanalysen bestätigt werden (siehe Abbildung 4.20). Es können jeweils Hauptkomponenten identifiziert werden, deren Scores gattungsspezifische Informationen enthalten: Für die Scores der ersten und vierten Hauptkomponenten der PCA von Daten der Ordnung *Coniferales* (Abbildung 4.20a) und für die Scores der ersten und fünften Hauptkomponente der PCA von Daten der Ordnung *Fagales* (Abbildung 4.20d) ist eine

Differenzierung einzelner Gattungen, beispielsweise von *Taxus*, *Pinus* und *Quercus* möglich. Bei den Scores der jeweils zweiten und dritten Hauptkomponenten (Abbildung 4.20b und e) sind allerdings gattungsspezifische Varianzen mit anderen Varianzen, die durch einzelne Spektren verursacht werden, überlagert. Dies ist z.B. daran erkennbar, dass ein einzelnes Spektrum der Gattung *Abies* einen sehr negativen Wert für die dritte Hauptkomponente aufweist (siehe Abbildung 4.20b), während alle anderen Spektren der Ordnung *Coniferales* weitestgehend durch Werte von ungefähr null charakterisiert sind. Betrachtet man die Scores der zweiten Hauptkomponente der PCA von *Coniferales*-Daten (siehe Abbildung 4.20b), so ist erkennbar, dass diese primär zwischen der bereits zuvor diskutierten Gruppierung aus *Picea*- und *Juniperus*-Spektren und den anderen Spektren der Ordnung *Coniferales* separiert. Die Loadings dieser Hauptkomponente (siehe Abbildung 4.20c, zweite Spur) zeigen Banden, die Adenin und Adeninderivaten zugeordnet werden können^{[100],[21]} (vgl. auch Tabelle 4.3). Da diese einen hohen Streuquerschnitt und eine hohe Affinität zu Metallnanopartikeln aufweisen,^[101] ist es wahrscheinlich, dass bei der Messung dieser Proben experiment-spezifische Varianzen in den Spektren auftreten, welche jedoch durch diese Hauptkomponente von den taxonomisch relevanten Varianzen separiert werden können. Betrachtet man die Loadings der anderen Hauptkomponenten (Abbildung 4.20c und f), so ist erkennbar, dass bei diesen eine Vielzahl unterschiedlicher Banden die Hauptkomponenten beeinflussen und dass hier eine Unterscheidung anhand spektraler Fingerabdrücke erfolgt. Dies wird durch die Transformierung zu einer Standardnormalverteilung erreicht.

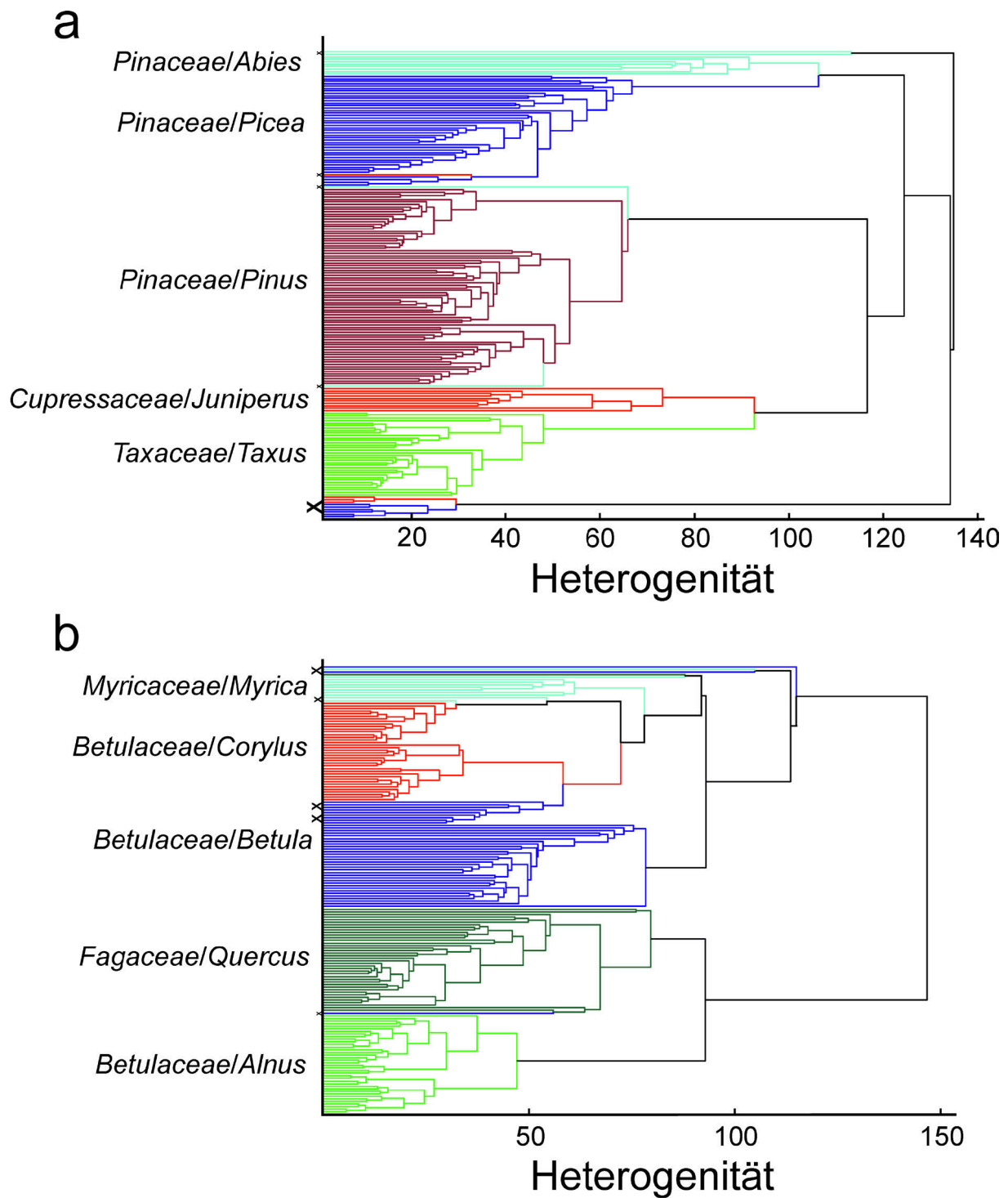


Abbildung 4.19: Ergebnis der Clusteranalyse von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a) und *Fagales* (b) als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden.

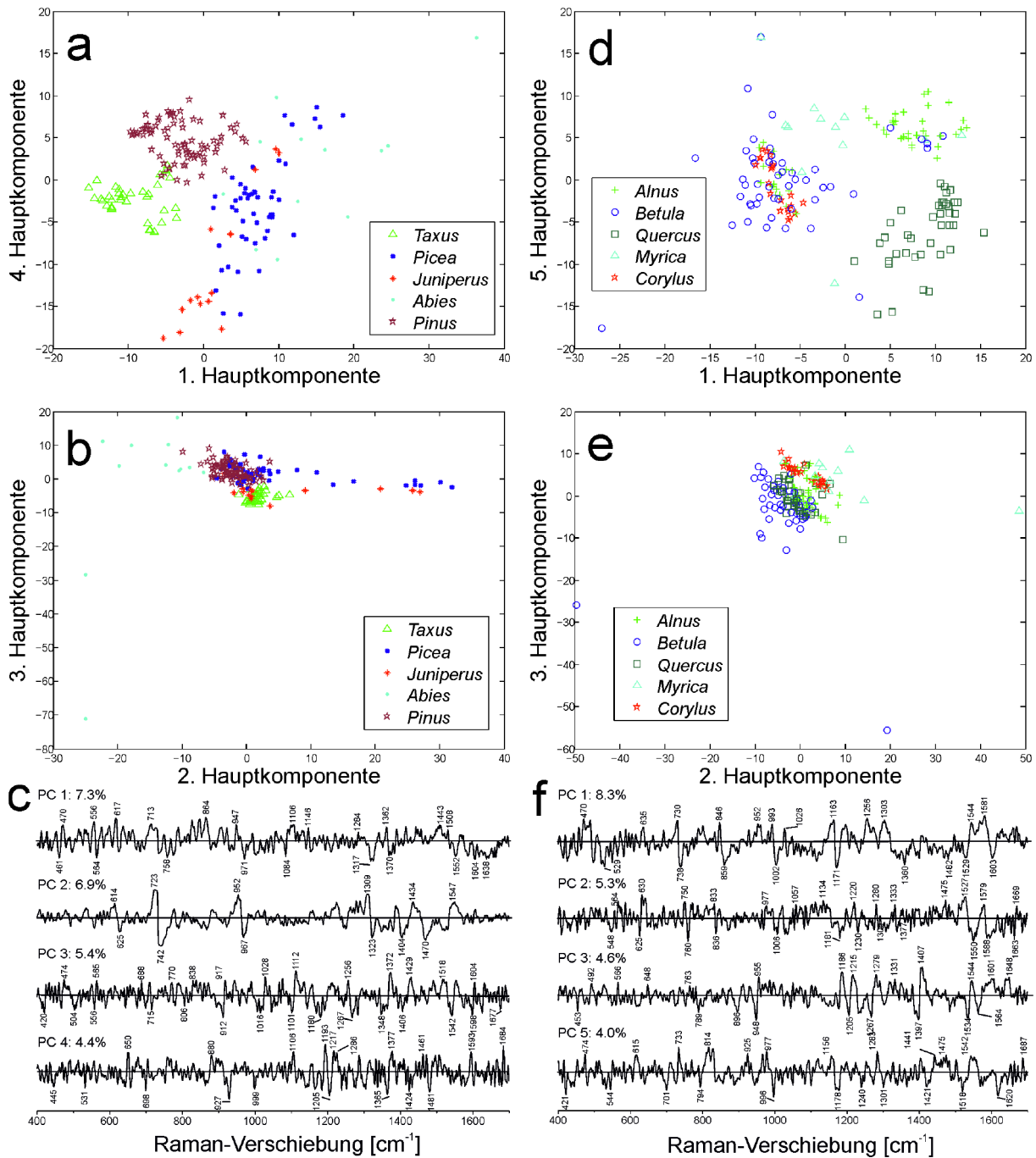


Abbildung 4.20: Ergebnisse der PCA von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a-c) und *Fagales* (d-f) als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Es sind die Scores der ersten, zweiten, dritten und vierten Hauptkomponente für Proben der Ordnung *Coniferales* (a+b), die Scores der ersten, zweiten, dritten und fünften Hauptkomponente für Proben der Ordnung *Fagales* (d+e), sowie die entsprechenden Loadings der jeweiligen Hauptkomponenten (c+e) gezeigt.

4.5.5 Untersuchung taxonomisch relevanter Daten der Gattungen

Äquivalent zur Klassifizierung auf der taxonomischen Ebene der Ordnung wurde anschließend eine Extraktion taxonomisch relevanter Daten aus den SERS-Datensätzen durchgeführt. Es wurden jeweils die Spektren gemittelt, die eine Zuordnung von mindestens 0.5 zu einer der jeweiligen Klassen aufwiesen. Abbildung 4.21 zeigt die Ergebnisse der HCA. Weder innerhalb der Ordnung *Coniferales*, noch der Ordnung *Fagales* ist eine klare Unterscheidung der Gattungen möglich. Allerdings werden jeweils einzelne Cluster gebildet, die eine hohe Anzahl von Spektren einzelner taxonomischer Gruppen enthalten. Bei der HCA von Daten der Ordnung *Coniferales* (Abbildung 4.21a) wird ein Cluster gebildet, der zu 80%

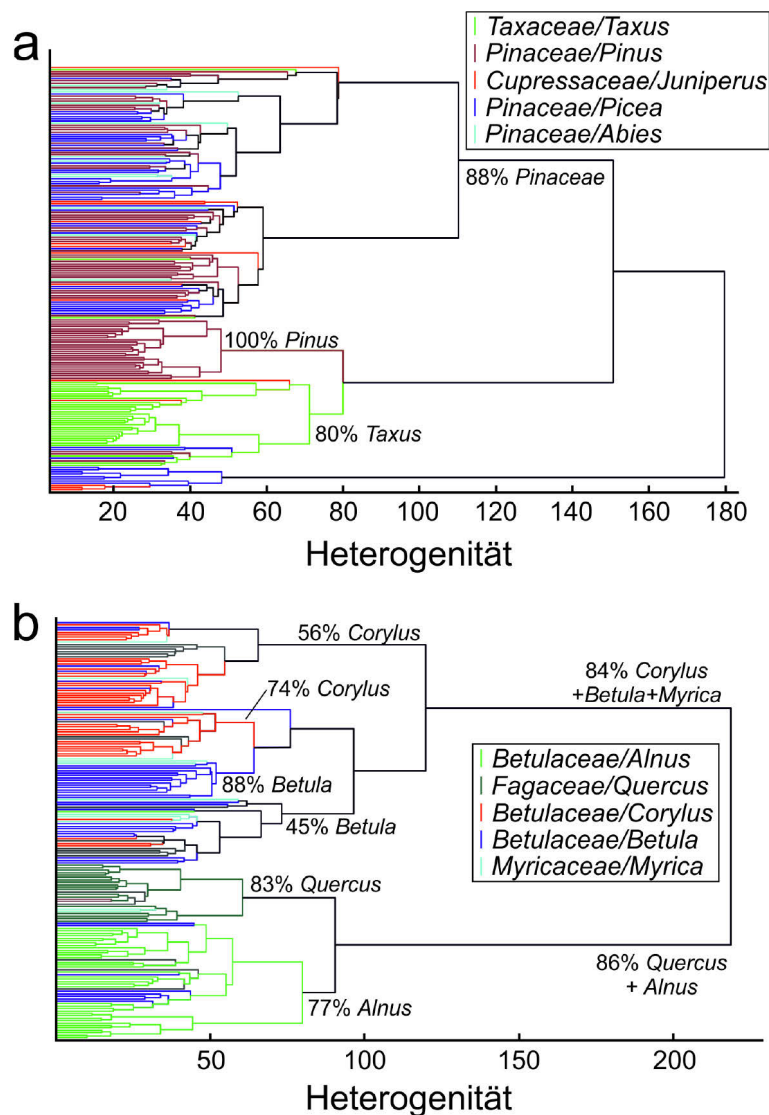


Abbildung 4.21: Ergebnis der Clusteranalyse von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a) und *Fagales* (b) als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden.

aus Spektren von Extrakten der Gattung *Taxus* besteht und den größten Teil der Datensätze dieser Gattung enthält. Eine Differenzierung vieler Datensätze zwischen denen der Familien *Taxaceae* und *Pinaceae* ist somit möglich. Des Weiteren wird ein Cluster gebildet, der durch gattungsspezifische Ähnlichkeiten charakterisiert ist und ausschließlich Datensätze der Gattung *Pinus* enthält (Abbildung 4.21a). Die Tatsache, dass in den Daten jedoch auch andere, gattungsspezifisch irrelevante Varianzen vorliegen, ist allerdings darin ersichtlich, dass sowohl die weiteren Datensätze dieser Gattung, als auch die von *Picea*-, *Abies*- und *Juniperus*-Pollenextrakten keine einzelnen Cluster bilden (Abbildung 4.21a), sondern Ähnlichkeiten zu den Daten anderer Gattungen zeigen.

Die größten spektralen Distanzen der Datensätze der *Fagales*-Pollenproben kann auf gattungsspezifische Unterschiede zwischen Extrakten von *Quercus* und *Alnus* und *Corylus*, *Betula* und *Myrica* zurückgeführt werden (Abbildung 4.21b). Innerhalb dieser Cluster sind Gruppierungen identifizierbar, die jeweils einen hohen Prozentsatz von Datensätzen von jeweils einer Gattung enthalten. Eine grobe Differenzierung zwischen Datensätzen verschiedener Gattungen der *Fagales*-Ordnung ist also anhand der HCA von taxonomisch relevanten Datensätzen möglich. Dass allerdings in den jeweiligen Clustern auch bis zu 55% an Datensätzen anderer Gattungen vorhanden sind (Abbildung 4.21b), zeigt, dass auch bei diesen Daten taxonomisch irrelevante Varianzen vorhanden sind, die eine deutlichere Unterscheidung der Datensätze verschiedener Gattungen verhindern. Dies ist auch daran ersichtlich, dass für die Datensätze der jeweiligen Gattungen, wie z.B. die von *Betula* und *Corylus* (Abbildung 4.21b, blaue und rote Cluster), jeweils mehrere Cluster gebildet werden, die eine hohe Heterogenität zueinander aufweisen.

Eine Möglichkeit, die verschiedenen Varianzen zu separieren, könnte die PCA der taxonomisch relevanten Daten bieten. Abbildung 4.22 zeigt das Ergebnis der PCA. Es ist deutlich erkennbar, dass anhand der ersten Hauptkomponente des *Coniferales*-Datensatzes (Abbildung 4.22a) eine Identifizierung einer Gruppierung von acht *Picea*- und drei *Juniperus*-Proben erfolgen kann. Diese konnte bereits bei der Untersuchung der taxonomisch spezifischen Daten (vgl. Abschnitt 4.5.4) aufgrund des hohen Einflusses von Banden, die Adenin und Adeninderivaten zugeordnet wurden, als nicht gattungsspezifische Varianz identifiziert werden. Die Loadings dieser Hauptkomponente (Abbildung 4.22c, erste Spur) zeigen, dass auch hier die Datensätze aufgrund dieser experimentell bedingten Varianzen separiert werden. Es sind allerdings auch gattungsspezifische Scores in den Hauptkomponenten detektierbar. So weist ein großer Teil der Datensätze der Gattung *Taxus* vermehrt positive Scores für die erste und vierte Hauptkomponente auf (Abbildung 4.22a und

b), während viele Datensätze der Gattung *Juniperus* durch negative Scores für die dritte und positive Scores für die vierte Hauptkomponente charakterisiert sind. Wie bei der Auswertung mit HCA (vgl. Abbildung 4.21) ist allerdings eine Unterscheidung von Datensätzen verschiedener Gattungen der Familie *Pinaceae* anhand der vorliegenden Daten nicht möglich. Auch bei der PCA von Datensätzen der Ordnung *Fagales* sind gattungsspezifische Varianzen erkennbar (Abbildung 4.21d bis e). Die bereits in der HCA als größte Varianz im Datensatz sichtbaren Unterschiede zwischen den taxonomischen Gruppen *Corylus*, *Betula* und *Myrica* und *Quercus* und *Alnus* (vgl. Abbildung 4.21) sind in der PCA in der ersten Hauptkomponente repräsentiert (Abbildung 4.22d). Weiterhin kann zwischen vielen Datensätzen der Gattungen *Alnus* und *Quercus* anhand der zweiten (Abbildung 4.22d), sowie anhand der dritten und vierten Hauptkomponente (Abbildung 4.22e) unterschieden werden. Allerdings ist auch bei dieser Auswertung ersichtlich, dass gattungsspezifische Varianzen durch andere Varianzen überlagert werden. Aus diesem Grund ist diese Klassifizierung nicht für alle Datensätze jeder Gattung möglich, und selbst eine grobe Differenzierung von Datensätzen der Gattungen *Betula*, *Corylus* und *Myrica* (Abbildung 4.21d und e) scheitert.

Aus der Auswertung taxonomisch relevanter Daten verschiedener Gattungen der Ordnungen *Coniferales* und *Fagales* kann zusammenfassend geschlossen werden, dass gattungsspezifische Varianzen in den Daten vorhanden sind, und teilweise einzelne taxonomische Gruppen von anderen separiert werden können. Diese Varianzen sind allerdings in hohem Maße von anderen Varianzen überlagert, so dass in vielen Fällen eine Differenzierung verschiedener taxonomischer Gruppen anhand von PCA und HCA nicht möglich ist. Die Gewichtung von Spektralbereichen der einzelnen SERS-Spektren durch das ANN ist somit für eine erfolgreiche Klassifizierung auf der taxonomischen Ebene der Gattung zwingend erforderlich.

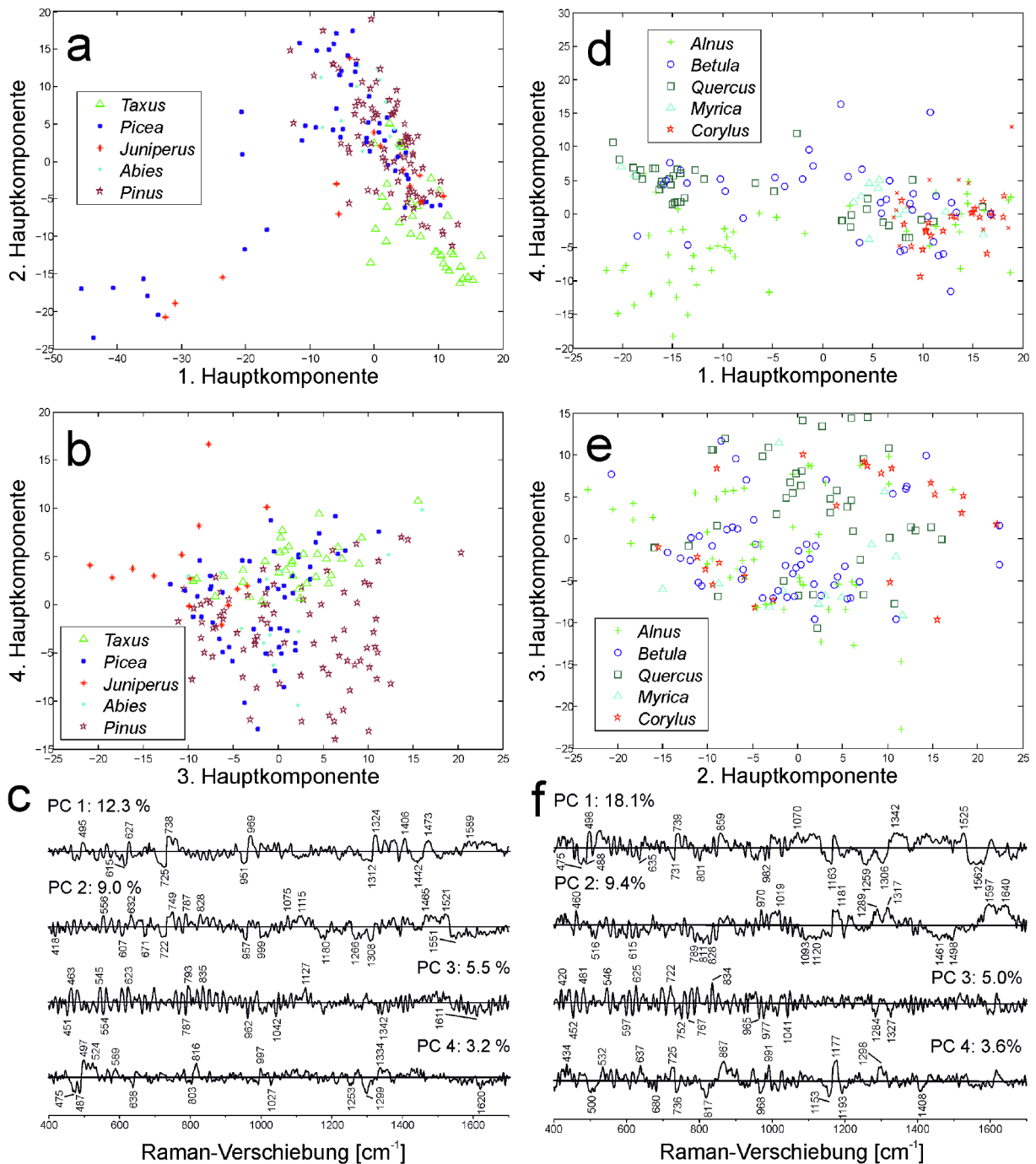


Abbildung 4.22: Ergebnisse der PCA von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a-c) und *Fagales* (d-f) als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Es sind die Scores der ersten, zweiten, dritten und vierten Hauptkomponente für Proben der Ordnung *Coniferales* (a+b), die Scores der ersten, zweiten, dritten und vierten Hauptkomponente für Proben der Ordnung *Fagales* (d+e), sowie die entsprechenden Loadings der jeweiligen Hauptkomponenten (c+f) gezeigt.

4.6 Schlussfolgerungen

Zusammenfassend kann man über die SERS-Untersuchung von wässrigen Pollenextrakten sagen, dass durch diese Methode die inneren, wasserlöslichen Bestandteile von Pollenkörnern analysiert werden können. Dadurch werden zu anderen schwingungsspektroskopischen Methoden komplementäre Informationen erhalten. Die SERS-Spektren können als spezifische Fingerabdrücke mit neuronalen Netzen ausgewertet und so eine Klassifizierung auf verschiedenen taxonomischen Ebenen erreicht werden. Damit wurde gezeigt, dass neuronale Netze, angewendet zur Analyse von SERS-Daten, nicht nur für die Identifizierung einzelner Bestandteile in Molekülgemischen^[81], sondern auch für die Klassifizierung komplexer, biologischer Extrakte nutzbar sind. Da mit SERS im Gegensatz zu Raman-^[13]- und FTIR-Untersuchungen^[16] die zellulären Bestandteile unabhängig von der Sporopollenin-Außenhülle analysiert werden können, ist diese Methode besonders vielversprechend, um auch die Physiologie von Pollen z.B. in Hinblick auf den Einfluss von Umweltfaktoren wie Virusinfektionen oder klimatischen Bedingungen zu untersuchen. Wie die hier diskutierten Ergebnisse zeigen, sind neuronale Netze auch für die Extraktion spezifischer Spektren aus den Datensätzen geeignet. Auf diese Weise konnte festgestellt werden, dass für die Unterscheidung der taxonomischen Ebene der Ordnung ganze Spektren spezifisch der Klassifizierung dienen können, während für die Unterscheidung der Gattungen eine Gewichtung einzelner Spektralbereiche der einzelnen SERS-Spektren durch das neuronale Netz für eine erfolgreiche Klassifizierung erforderlich ist.

5. MALDI-TOF MS zur taxonomischen Klassifizierung von Pollen

In den letzten Jahren und Jahrzehnten hat sich die *matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS) zu einer beliebten Methode zur Analyse von komplexen biologischen Proben entwickelt. Ein Hauptanwendungsgebiet liegt dabei in der Protein- und Proteomikuntersuchung für medizinische Anwendungen^[105]. So wurde u.a. der Einfluss von Leberzirrhose^[106], Diabetes^[107], Hepatitis^[108], Multiple Sklerose^[109], Tuberkulose^[110] oder verschiedener Herzerkrankungen^{[111],[112]} auf die Proteinzusammensetzung des Körpers erforscht. Da die Untersuchung von Tumoren zurzeit von besonderem Interesse ist, wurde MALDI-TOF MS ebenfalls erfolgreich für die Charakterisierung^[113] und Identifizierung^{[114],[115]} von Krebszellen eingesetzt. Auch in diesem Zusammenhang zeigte sich, dass MALDI-TOF MS für die automatische Aufnahme spezifischer Spektren, die als Fingerabdrücke dienen, verwendet werden kann.^[116] Diese können anschließend mit multivariaten Methoden kombiniert werden, um die Blutproben von krebserkrankten und von gesunden Patienten zu unterscheiden.^[117] Die Verknüpfung von MALDI-TOF MS-Daten mit multivariaten Verfahren erwies sich auch für die Entwicklung bildgebender Verfahren zur Gewebeuntersuchung^[61] als sinnvoll, beispielsweise um den Einfluss einer Alzheimer Erkrankung auf das Gehirn zu analysieren^[118].

Trotz der vielversprechenden Ergebnisse dieser Methode in der medizinischen Forschung, wurden pflanzliche Proben bisher relativ selten untersucht. Zumeist wurden dabei lediglich einzelne molekulare Bestandteile, wie zum Beispiel Carotinoide^[119] qualitativ und quantitativ erfasst. Auch die Untersuchung von Pollen mit MALDI-TOF MS ist bis jetzt weitestgehend auf die Analyse einzelner Komponenten, wie Sporopollenin^[120] oder verschiedener Allergene beschränkt. Es wurden beispielsweise die Pollenallergene der Olive^{[121],[122]}, Tomate^[123], Haselnuss^{[124],[125]}, Zypresse^[126], des Beifußes^[127] und verschiedener Gräser^[128] charakterisiert. In diesem Kapitel werden massenspektrometrische Fingerabdrücke von Pollenkörnern, die sich aus der Überlagerung von Signalen von einer Vielzahl an Molekülen ergeben, diskutiert, und ihre Eignung für die Klassifizierung von Pollen illustriert. Ähnliche Anwendungen wurden bereits für die Unterscheidung verschiedener Pilzbestandteile^[129], sowie für die taxonomische Klassifizierung von Bakterien^{[63],[130],[131]} und Pilzen^[62] entwickelt. In diesem Zusammenhang wurde gezeigt, dass multivariate Methoden, kombiniert mit massenspektrometrischen Daten, taxonomische Verwandtschaftsverhältnisse abbilden können.^[28] Dabei können auch die für die Klassifizierung relevanten Massenbereichen in den

Spektren identifiziert werden.^[132] Kürzlich wurden mit dieser leistungsstarken Kombination sogar ganze, komplexe Organismen untersucht, um verschiedene Zeckenarten zu identifizieren.^[64] Dass ähnliche Anwendungen auch für pflanzliche Proben möglich sind, wurde durch die Klassifizierung von Blättern unterschiedlicher Verwandtschaftsverhältnisse demonstriert.^[133]

Um eine ähnliche Methode für die Klassifizierung von Pollen zu entwickeln, sollen im Folgenden zunächst kommerziell erhältliche Pollenproben untersucht und dabei die Reproduzierbarkeit und Spezifität der Methode überprüft werden (Abschnitt 5.1). Anschließend werden Pollenproben aus dem Feld analysiert, um zu prüfen, inwiefern taxonomische Verwandtschaftsverhältnisse aus den MALDI-TOF MS-Daten reproduziert werden können (Abschnitt 5.2).

5.1 Eignung von MALDI-TOF MS zur Klassifizierung von Pollenproben

Um MALDI-TOF MS für die taxonomische Klassifizierung von Pollen verwenden zu können, muss untersucht werden, ob diese Methode reproduzierbare, taxonomisch spezifische Fingerabdrücke liefert. Es wurden dazu kommerziell erhältliche Pollenproben (siehe Tabelle 3.1) untersucht, da bei diesen Pollenkörner in sehr großen Mengen vorhanden waren und dieselbe Probe demnach beliebig oft mit MALDI-TOF MS untersucht werden konnte. Dabei wurden Pollenproben sehr unterschiedlicher taxonomischer Verwandtschaft einschließlich zweier verschiedener Chargen der Art *Betula fontinalis* verwendet. Abbildung 5.1 zeigt die Spektren der verschiedenen Pollenproben. Alle Spektren weisen Banden im Massenbereich zwischen m/z 1000 und 20000 auf, die möglicherweise als spezifische Fingerabdrücke dienen könnten. MS/MS Untersuchungen zeigen, dass diverse Banden auf die für Glykoproteine typischen Fragmente von Oligosacchariden, N-Acetylhexosamin und Sialinsäure zurückgeführt werden können.^[134] Die für Hexoseeinheiten typischen Bandenabstände von 162 Da sind auch im Spektrum von *Juniperus scopulorum* erkennbar (Abbildung 5.1, fünfte Spur rechts). Die beiden Spektren von unterschiedlichen Proben der Art *Betula fontinalis*, sind fast identisch (Abbildung 5.1, vierte und fünfte Spur links). Dies spricht dafür, dass durch die Untersuchung ein spezifischer, taxonomischer Fingerabdruck der jeweiligen Pollenprobe erhalten werden kann. Die große Ähnlichkeit der beiden Spektren der Gattung *Artemisia* (Abbildung 5.1, zweite und dritte Spur links), sowie der verschiedenen Arten der Gattung *Populus* (Abbildung 5.1, erste bis dritte Spur rechts), legt die Vermutung nahe, dass spezifische Fingerabdrücke erhalten werden, die Ähnlichkeiten in taxonomisch verwandten Pollenproben aufweisen. Um diese Vermutungen zu prüfen, wurde das unüberwachte, multi-

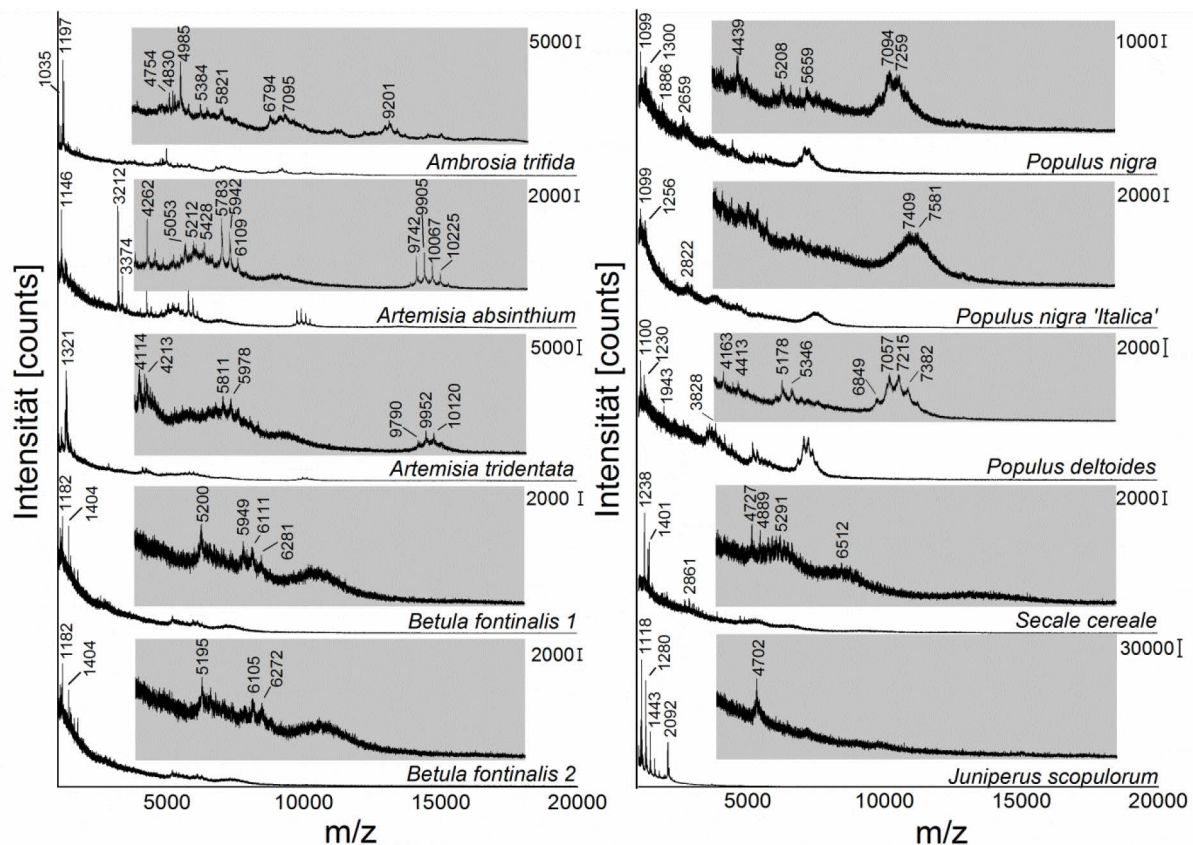


Abbildung 5.1: MALDI-TOF MS-Spektren kommerziell erhältlicher Pollenproben im Bereich zwischen m/z 1000 und 20000, bzw. zwischen m/z 4000 und 11000 (vergrößerter Bereich in grau).

variate Verfahren der hierarchischen Clusteranalyse (HCA) angewendet. Dazu wurden für jede der zehn Pollenproben 24 MALDI-TOF MS-Spektren aufgenommen und ausgewertet. Das Ergebnis der HCA dieser Spektren (Abbildung 5.2) zeigt, dass neun Cluster gebildet wurden, die jeweils die Spektren einer Pollenart enthalten. Dass davon keine Abweichung auftritt, beweist, dass es sich um sehr spezifische und reproduzierbare Fingerabdruckspektren handelt. Da ein Cluster vorhanden ist (Abbildung 5.2, vierter Cluster von unten), der alle Spektren beider Chargen der Art *Betula fontinalis* enthält, zwischen welchen anhand dieser Analyse nicht differenziert werden kann, zeigt weiterhin, dass dieser spezifische Fingerabdruck auch taxonomisch relevant ist. Dieses Spektrum kann demnach in beinahe identischer Form erhalten werden, wenn eine Pollenprobe der gleichen Art untersucht wird. Die Tatsache, dass die Cluster der Pollenspektren von *Artemisia tridentata* und *Artemisia absinthium* (Abbildung 5.2, unten), sowie *Populus deltoides* und *Populus nigra* (Abbildung 5.2, zweiter und dritter Cluster von oben), größere Ähnlichkeiten zueinander aufweisen, als zu den Clustern anderer Pollenarten, unterstützt die Vermutung, dass auch taxonomische Verwandtschaftsverhältnisse aus den MALDI-TOF MS-Daten reproduziert werden können.

Ob dies möglich ist, soll im nun folgenden Abschnitt systematisch anhand von Pollenproben aus dem Feld analysiert werden.

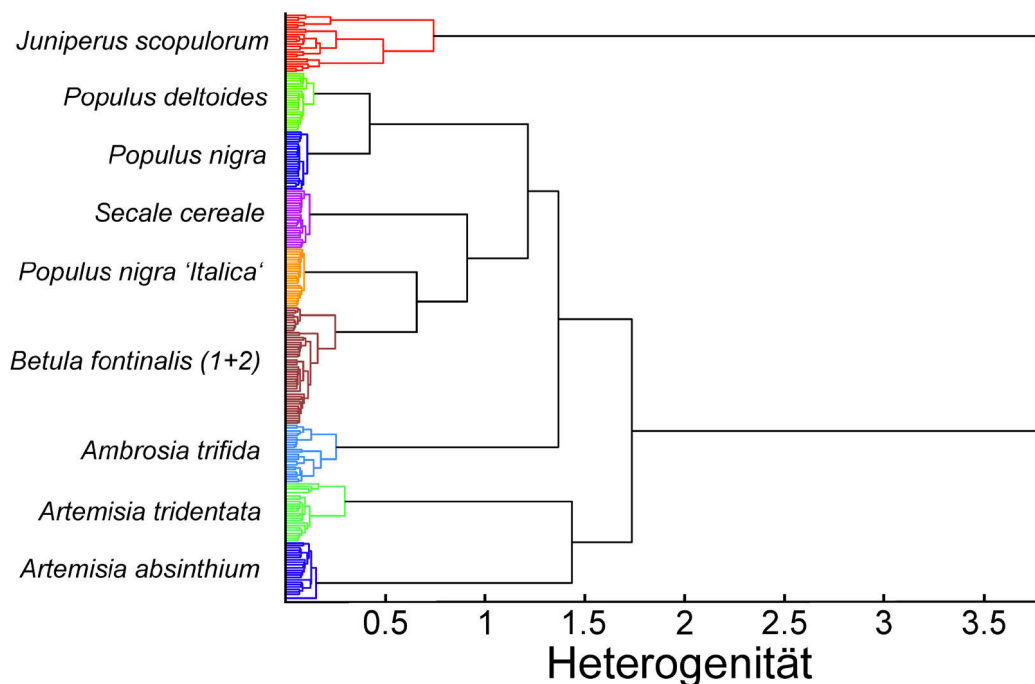


Abbildung 5.2: Ergebnis der Klassifizierung von 240 Spektren der 10 kommerziell erhältlichen Pollenproben mit HCA. Es wurden euklidische Distanzen im Massebereich zwischen m/z 1000 und 15000 und der Ward's Algorithmus verwendet.

5.2 Wiedergabe taxonomischer Verwandtschaftsverhältnisse aus MALDI-TOF MS-Daten

Anhand der Untersuchung kommerziell erhältlicher Pollenproben wurde gezeigt, dass durch MALDI-TOF MS reproduzierbare Fingerabdruckspektren der Pollenkörner erhalten werden können. Diese enthalten spezifische taxonomische Informationen und können mit multivariaten Verfahren klassifiziert werden. Um zu untersuchen, inwiefern sich taxonomische Verwandtschaftsverhältnisse aus den MALDI-TOF MS-Daten reproduzieren lassen, wurden Pollenproben analysiert, die in Berliner Parks und im Botanischen Garten Berlin-Dahlem gesammelt wurden. Es wurden dabei Pollenproben von 74 Arten und elf Gattungen in sechs Familien der Ordnungen *Fagales* und *Coniferales* (siehe Tabelle 3.2) untersucht. Aufgrund der geringeren Probenmenge wurde im Gegensatz zur Untersuchung der kommerziell erhältlichen Pollenproben die Methode der schnelleren, lösungsmittelfreien Präparation verwendet.^[134] Abbildung 5.3 zeigt exemplarisch Spektren verschiedener Arten. Die Spektren der Ordnung *Coniferales* weisen hauptsächlich Banden auf, die sich unterhalb

von m/z 5000 befinden. Dabei treten des Öfteren, besonders jedoch in Spektren der Gattung *Juniperus*, charakteristische Distanzen von 162 Da auf. Wie bereits zuvor diskutiert, handelt es sich bei diesen um typische Fragmente von Oligosacchariden^[134] (siehe auch Abschnitt 5.1). In den Spektren der Ordnung *Fagales* treten zusätzlich zu Massen unterhalb von m/z 5000 auch Massen auf, die sich zwischen m/z 5000 und m/z 10000 befinden. Es ist offensichtlich, dass Spektren der Ordnung *Fagales* durch diese spektralen Merkmale relativ einfach von Spektren der Ordnung *Coniferales* unterschieden werden können.

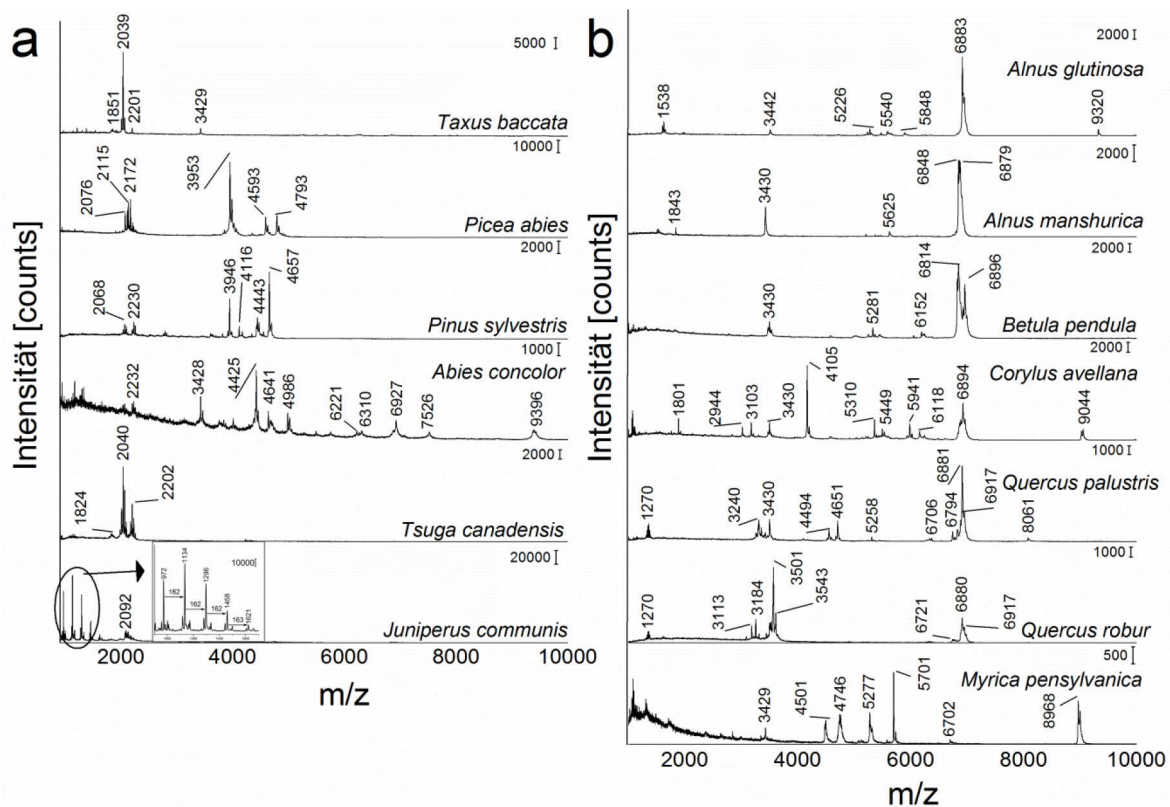


Abbildung 5.3: Charakteristische MALDI-TOF MS-Spektren ausgewählter Pollenproben der Ordnungen *Coniferales* (a) und *Fagales* (b).

5.2.1 HCA von MALDI-TOF MS-Spektren

Um die Möglichkeit einer unüberwachten, taxonomischen Klassifizierung durch Verwendung von MALDI-TOF MS-Daten zu testen, wurde anschließend eine HCA mit den Daten durchgeführt. Das Ergebnis ist in Abbildung 5.4 gezeigt. Es werden zwei Cluster gebildet, die jeweils v.a. die Spektren einer Ordnung enthalten. Es können demnach die größten Unterschiede im Datensatz auf Unterschiede von Spektren verschiedener taxonomischer Ordnungen zurückgeführt werden, und 95% der Spektren befinden sich im Cluster der jeweils korrekten Ordnung. Bei 6 der 9 falsch klassifizierten Pollenspektren handelt es sich um die

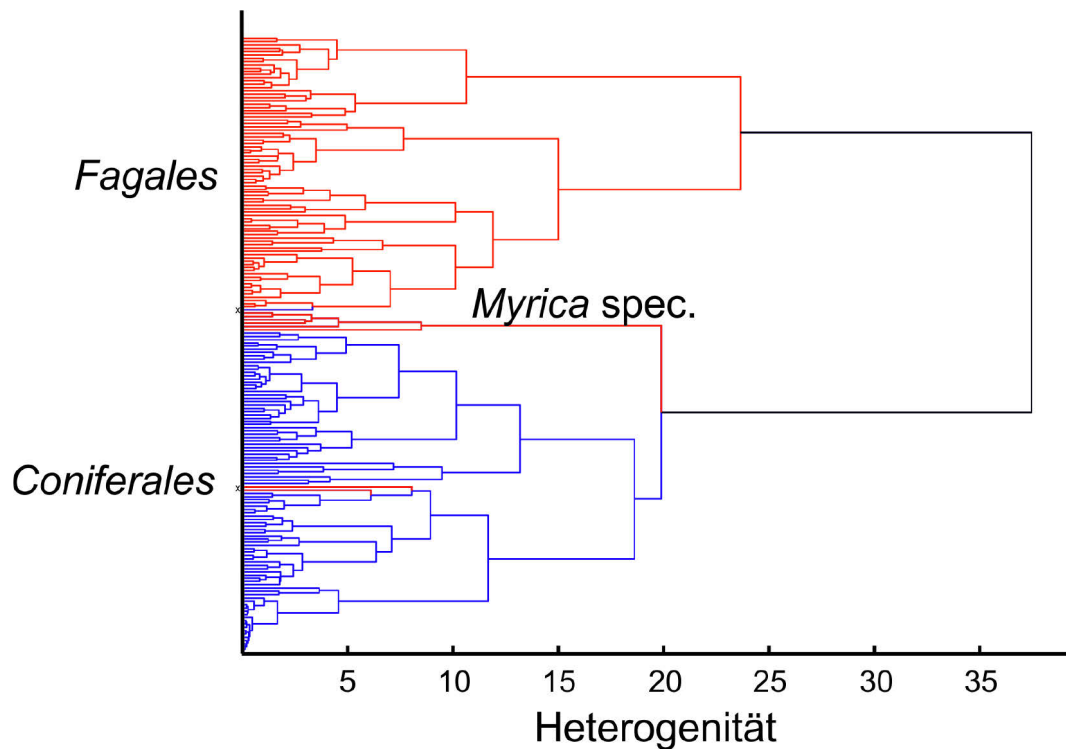


Abbildung 5.4: Ergebnis der Klassifizierung von 188 Pollenspektren von 74 verschiedenen Arten der Ordnungen *Coniferales* und *Fagales* mit HCA. Es wurden euklidische Distanzen im Massebereich zwischen 1000 und 10000 m/z und der Ward's Algorithmus verwendet.

Spektren der Gattung *Myrica*, die einen separaten Cluster bilden. Dieser weist hohe Heterogenität sowohl zum Cluster der Ordnung *Fagales*, als auch zu dem der Ordnung *Coniferales* auf. Da es sich bei *Myrica* um eine Gattung handelt, die sich auch im Phänotyp sehr stark von den anderen untersuchten Proben von *Fagales* unterscheidet, ist es nicht überraschend, dass auch die Massenspektren der Pollenproben große Unterschiede aufweisen (siehe Abbildung 5.3, zweite Spur rechts). Definiert man die Gattung *Myrica* als eigenständige taxonomische Gruppe, so werden 98% der Pollenproben der jeweils korrekten Ordnung zugewiesen (siehe Abbildung 5.4). Lediglich zwei Spektren von *Fagales* und ein Spektrum von *Coniferales* befinden sich im Cluster der jeweils anderen Ordnung. Betrachtet man die zugehörigen Massenspektren und vergleicht diese mit Spektren der jeweils gleichen Gattung (siehe Abbildung 5.5), so ist erkennbar woraus diese falsche Zuordnung resultiert. Im Falle der beiden *Fagales*-Spektren, bei denen es sich um Spektren der Gattung *Alnus* handelt, treten zusätzliche, intensive Banden in verschiedenen Massenbereichen auf (siehe Abbildung 5.5a bis c), die auf zusätzliche probenspezifische Bestandteile hinweisen. Die falsch zugeordnete *Coniferales*-Pollenprobe betrifft das Spektrum einer Probe der Gattung *Abies*. (siehe Abbildung 5.5d bis e) Die fehlerhafte Klassifizierung ist in diesem Fall auf große Intensitätsunterschiede der für diese Gattung spezifischen Banden zurückzuführen.

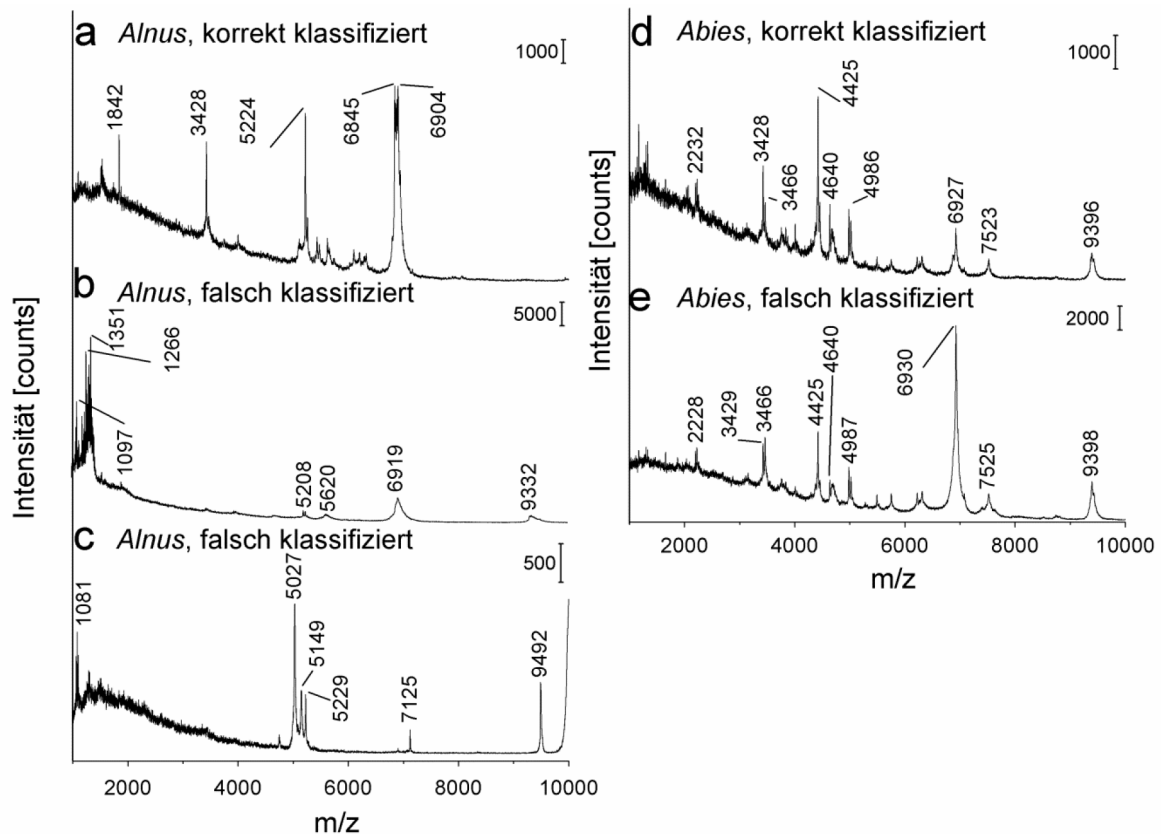


Abbildung 5.5: MALDI-TOF MS-Spektren von *Alnus firma* (a), *Alnus cordata* (b), *Alnus subcordata* (c), und *Abies concolor* (d+e).

Um zu überprüfen, inwiefern taxonomische Verwandtschaftsverhältnisse der Familien, Gattungen und Arten aus den MALDI-TOF MS-Daten abgebildet werden können, wurden einzelne Clusteranalysen der *Fagales*- und *Coniferales*-Spektren durchgeführt (Abbildung 5.6). Die größte Heterogenität innerhalb der *Coniferales*-Proben kann auf Unterschiede der beiden Gruppen aus *Taxus*, *Tsuga*, *Picea* und *Juniperus* und *Pinus* und *Abies* zurückgeführt werden (Abbildung 5.6a). Beide Cluster können weiterhin in kleinere Cluster unterteilt werden, die die Spektren der jeweiligen Gattungen enthalten. Innerhalb dieser Cluster sind Spektren vorhanden, die hohe Ähnlichkeiten innerhalb ihrer Art aufweisen. Dies gilt für Spektren von *Abies concolor*, *Juniperus rigida*, *Juniperus communis*, *Picea breweriana*, *Picea purpurea*, *Picea abies*, *Pinus mugo* und *Pinus sylvestris*. Betrachtet man die taxonomische Ebene der Familien, so fällt auf, dass *Pinaceae*-Spektren sowohl den Spektren der *Taxaceae* als auch denen der *Cupressaceae* ähneln (siehe Abbildung 5.6a). Eine Unterscheidung verschiedener Familien von *Coniferales*-Pollen ist demnach anhand von MALDI-TOF MS-Daten und HCA nicht möglich. Dass diese Abstufung mit FTIR-Untersuchungen erreicht werden kann,^[16] zeigt, dass durch MALDI-TOF MS andere chemische Bestandteile der Pollenkörner analysiert werden können, als durch schwingungsspektroskopische Methoden.

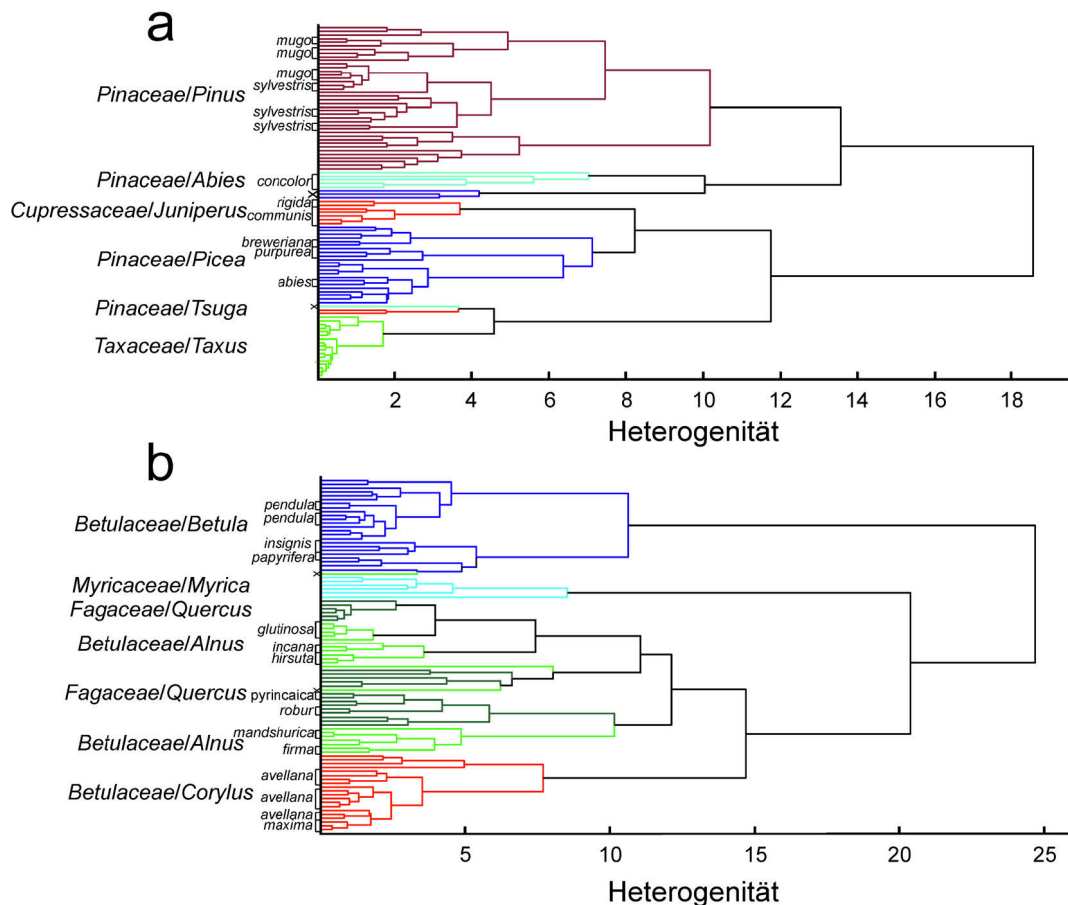


Abbildung 5.6: Ergebnis der HCA von 97 MALDI-TOF MS-Pollenspektren der Ordnung *Coniferales* (a) und von 91 MALDI-TOF MS-Spektren der Ordnung *Fagales* (b). Es wurden euklidische Distanzen im Massebereich zwischen 1000 und 10000 m/z und der Ward's Algorithmus verwendet.

Betrachtet man die Ergebnisse der Clusteranalyse der *Fagales*-Spektren (Abbildung 5.6b), so ist erkennbar, dass auch bei dieser Ordnung gattungsspezifische Cluster auftreten: Es wird jeweils ein Cluster für *Betula*, *Myrica* und *Corylus* gebildet. Auch bei diesen Gattungen weisen die Spektren einzelner Arten Ähnlichkeiten auf. Dies betrifft Spektren von *Betula pendula*, *Betula insignis*, *Corylus avellana* und *Corylus maxima*. Die Spektren der Gattungen *Alnus* und *Quercus* zeigen jedoch eine hohe spektrale Varianz und bilden verschiedene Cluster, die nicht eindeutig von den Clustern der jeweils anderen Gattung unterschieden werden können. Dies lässt darauf schließen, dass zwischen diesen beiden Gattungen mit MALDI-TOF MS in Kombination mit HCA nicht differenziert werden kann. Die Ursache dafür ist ersichtlich, wenn man die Spektren von *Alnus glutinosa* und *Alnus mandshurica*, bzw. die von *Quercus palustris* und *Quercus robur* (siehe Abbildung 5.3), die sich jeweils in unterschiedlichen Cluster befinden, vergleicht. Diese enthalten jeweils die gleichen Banden, die aber unterschiedliche Intensitäten besitzen. Da große Unterschiede der Bandenintensitäten auch als Grund dafür identifiziert wurde, dass eine *Abies*-Pollenprobe zur falschen Ordnung

gruppiert wurde (siehe Abbildung 5.5), ist es sinnvoll, eine Separierung unterschiedlicher Varianzen im Datensatz anzustreben anstatt spektrale Distanzen zu vergleichen.

5.2.2 PCA von MALDI-TOF MS-Spektren

Um die Auswertung auf die Varianzen im Datensatz auszurichten und diese Varianzen auf spezifische Masse-zu-Ladungs-Verhältnisse zurückführen zu können, wurde eine Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) der MALDI-TOF MS-Daten durchgeführt. Die Scores und Loadings sind in Abbildung 5.7 gezeigt. Die meisten Spektren der beiden Ordnungen können anhand der ersten Hauptkomponente, die 27.3% der Varianz im Datensatz repräsentiert, differenziert werden (Abbildung 5.7a). Die entsprechenden Loadings (Abbildung 5.7b) zeigen, dass die Spektren der Ordnung *Coniferales* positive Werte für die erste Hauptkomponente aufweisen (Abbildung 5.7a) und mit Banden unter m/z 5000 korrelieren. Im Gegensatz dazu korrelieren negative Scores, die charakteristisch für *Fagales*-Spektren sind (Abbildung 5.7a), v.a. mit Banden über m/z 5000 (Abbildung 5.7b). Dies bestätigt die Schlussfolgerungen, die bereits bei der Betrachtung der Spektren getroffen wurden (siehe Beginn von Abschnitt 5.2) und die Unterscheidung der Ordnungen entspricht den Ergebnissen der HCA (Abbildung 5.4). Um zu untersuchen, ob taxonomische Verwandtschaftsverhältnisse der Familien und Gattungen aus den MALDI-TOF MS-Daten reproduziert werden können, wurden, analog zur HCA (Abbildung 5.6), zwei von einander unabhängigen Hauptkomponentenanalysen der Spektren der Ordnungen *Coniferales* und *Fagales* durchgeführt (Abbildung 5.8). Die Ergebnisse der PCA der Pollenspektren der

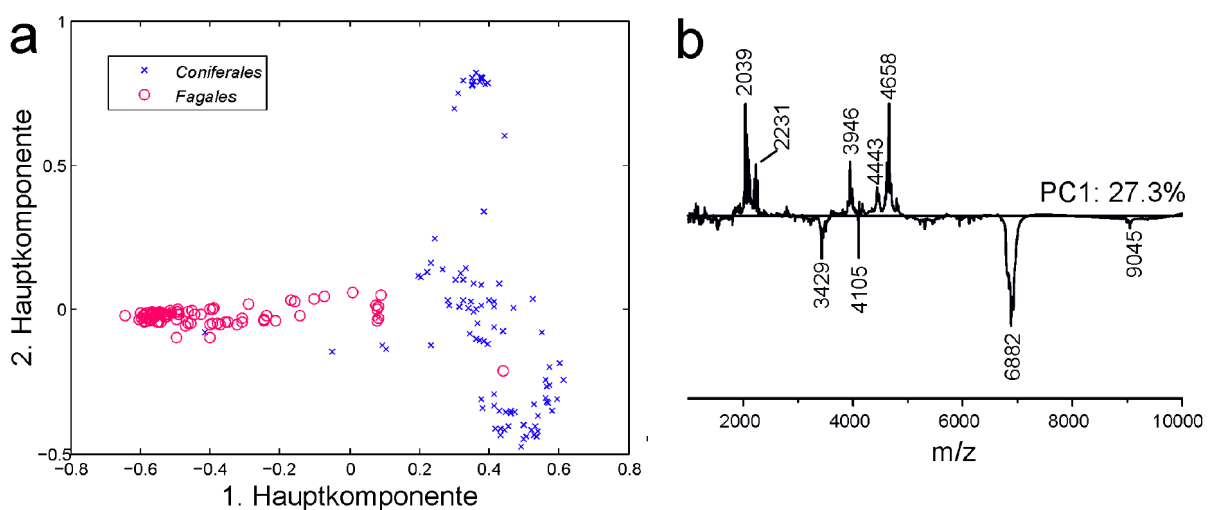


Abbildung 5.7: Ergebnis der PCA von 188 MALDI-TOF MS-Pollenspektren der Ordnungen *Coniferales* und *Fagales*. Es sind die Scores der ersten beiden Hauptkomponenten (a) und die Loadings und Varianz der ersten Hauptkomponente (b) gezeigt.

Ordnung *Coniferales* (Abbildung 5.8a bis c) demonstrieren, dass die erste, zweite und vierte Hauptkomponente eine Unterscheidung zwischen den verschiedenen Gattungen ermöglicht. *Taxus*- und *Tsuga*-Pollenspektren werden dabei vor allem durch positive Scores der ersten Hauptkomponente (Abbildung 5.8a und b), die mit einer Bande bei m/z 2039 korrelieren (Abbildung 5.8c, erste Spur), beschrieben. Positive Scores für die zweite Hauptkomponente sind charakteristisch für *Picea*-Pollenspektren (Abbildung 5.8b). Anhand der entsprechenden Loadings (Abbildung 5.8c, zweite Spur) ist sichtbar, dass diese mit Banden bei m/z 2077, 2115, 3955, 4594 und 4795 korrelieren. *Pinus*-Pollenspektren weisen negative Werte für die erste und zweite Hauptkomponente auf, was auf Banden bei m/z 3947 und 4657 zurückgeführt werden kann (Abbildung 5.8c, erste und zweite Spur). Charakteristisch für Pollenspektren der Gattung *Juniperus* sind Banden der Oligosaccharide (siehe auch Abbildung 5.3), die sich in positiven Loadings der vierten Hauptkomponente wiederfinden (Abbildung 5.8c, dritte Spur) und deshalb mit positiven Scores dieser Hauptkomponente korrelieren. *Abies*-Pollenspektren sind in der mit diesem Datensatz durchgeführten PCA durch Scores aller Hauptkomponenten in der Nähe von null charakterisiert (Abbildung 5.8a und b). Die Scores der Spektren von *Pinus*, *Picea*, *Abies* und *Tsuga*, bei denen es sich um Gattungen der *Pinaceae*-Familie handelt, sind sich nicht besonders ähnlich. Dies zeigt, dass eine Klassifizierung auf Ebene der Familien auch mit varianzgewichteter PCA nicht möglich ist und bestätigt, dass mit MALDI-TOF MS andere Bestandteile der Pollenchemie analysiert werden, als bei schwingungsspektroskopischen Methoden^[16] (siehe auch Abbildung 5.6).

Auch die verschiedenen Gattungen der Ordnung *Fagales* können durch die Scores der Hauptkomponenten der entsprechenden PCA (Abbildung 5.8d und e) unterschieden werden. Charakteristisch für Spektren der Gattung *Betula* ist dabei die Bande bei m/z 6814, die die negativen Loadings der ersten beiden Hauptkomponenten in hohem Masse beeinflusst (Abbildung 5.8f, erste und zweite Spur). Deshalb weisen auch die entsprechenden Scores negative Werte für Pollenspektren dieser Gattung auf (Abbildung 5.8d). Für *Corylus*-Pollenspektren werden positive Werte für die zweite und v.a. die dritte Hauptkomponente gebildet (Abbildung 5.8e), was in positiven Loadings dieser Hauptkomponenten für die gattungsspezifischen Banden bei m/z 4105, 5310, 5941, 6118 und 9045 resultiert (Abbildung 5.8f, zweite und dritte Spur). *Myrica*-Pollenspektren weisen positive Scores für die erste und negative Scores für die zweite Hauptkomponente auf (Abbildung 5.8d), weshalb die spezifischen Banden dieser Gattung, u.a. bei m/z 5701 und 8969, auch in den entsprechenden Loadings der Hauptkomponenten auftreten (Abbildung 5.8f, erste und zweite Spur). Die Scores der ersten beiden Hauptkomponenten sind für die Gattungen *Alnus* und *Quercus* kaum

unterscheidbar (Abbildung 5.8d). Dies bestätigt die Ergebnisse der HCA (siehe Abbildung 5.6b), bei der zwischen den Spektren dieser beiden Gattungen nicht differenziert werden kann. Betrachtet man allerdings die Scores der dritten Hauptkomponente, so ist erkennbar, dass die meisten Spektren der beiden Gattungen durch diese Hauptkomponente unterschieden werden können. *Quercus*-Pollenspektren sind dabei durch negative Scores cha-

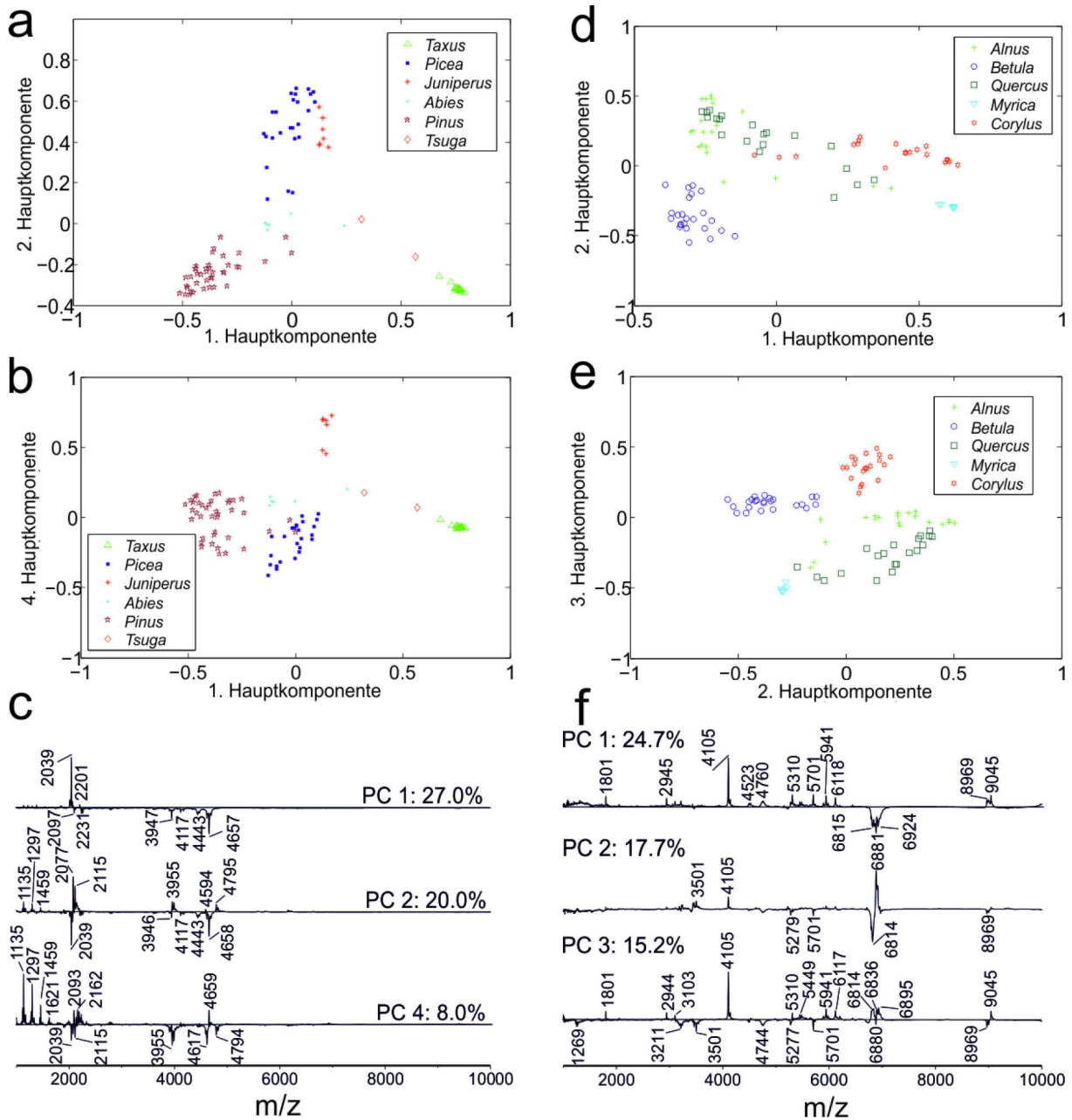


Abbildung 5.8: Ergebnisse der Hauptkomponentenanalysen von 97 MALDI-TOF MS-Pollenspektren der Ordnung Coniferales (a-c) und von 91 MALDI-TOF MS-Pollenspektren der Ordnung Fagales (d-f). Es sind die Scores der ersten, zweiten und vierten Hauptkomponente für Proben der Ordnung Coniferales (a+b), die Scores der ersten, zweiten und dritten Hauptkomponente für Proben der Ordnung Fagales (d+e), sowie die entsprechenden Loadings und Varianzen der jeweiligen Hauptkomponenten (c+e) gezeigt.

rakterisiert (Quadrate in Abbildung 5.8e), während Spektren der Gattung *Alnus* durch Werte in der Nähe von null beschrieben werden (Kreuze in Abbildung 5.8e). Deshalb befinden sich charakteristische Banden von *Quercus* in den negativen Loadings dieser Hauptkomponenten, z.B. bei m/z 1269 und 6880 (Abbildung 5.8f, dritte Spur). Wie bei der Untersuchung der Spektren der Ordnung *Coniferales* sind auch bei den Spektren von *Fagales* keine Ähnlichkeiten zwischen den Spektren der jeweiligen Familien erkennbar. Dies wird dadurch deutlich, dass Spektren der Gattungen *Betula*, *Alnus* und *Corylus*, die alle Bestandteil der *Betulaceae*-Familie sind, keine besonders ähnlichen Scores aufweisen.

Ob MALDI-TOF MS-Daten für die Unterscheidung verschiedener Arten einer Gattung verwendet werden können, wurde in voneinander unabhängigen Hauptkomponentenanalysen der Spektren der jeweiligen Gattungen überprüft (Abbildung 5.9 und Abbildung 5.10). Bei der Untersuchung der Spektren der Ordnung *Coniferales* ist ersichtlich, dass anhand der Scores zwischen den meisten Arten von *Picea* (Abbildung 5.9a), *Abies* (Abbildung 5.9b) und *Juniperus* (Abbildung 5.9c) und manchen Arten von *Pinus* (Abbildung 5.9d) und *Taxus* (Abbildung 5.9e) differenziert werden kann. Die Loadings der jeweiligen Hauptkomponenten (Abbildung 5.9f-j) sind auf die als gattungsspezifisch definierten Banden (siehe vorigen Abschnitt) zurückführbar. Sie repräsentieren damit artspezifische, qualitative Merkmale, die sich zwischen den Arten, z.B. durch verschiedene relative Intensitäten unterscheiden. So beeinflussen z.B. die Banden bei m/z 2077, 2115, 3953, 4593 und 4793 die PCA der Spektren der Gattung *Picea* (Abbildung 5.9b), die charakteristischen Banden der Oligosaccharide bei m/z 1133, 1297 und 1458 (vgl. Abbildung 5.3), sowie Banden bei m/z 2077 und 2162 die PCA der Spektren von *Juniperus* (Abbildung 5.9h) und die Bande bei m/z 2041 die PCA der Spektren von *Taxus*. Auch die Varianzen der PCA der Spektren der Gattung *Pinus* deuten auf Unterschiede der gattungsspezifischen Banden, u.a. bei m/z 3947 und 4659 hin.

Durch die Scores der einzelnen Hauptkomponentenanalysen der Ordnung *Fagales* kann ebenfalls zwischen den meisten Arten von *Alnus* (Abbildung 5.10a), *Betula* (Abbildung 5.10b), *Corylus* (Abbildung 5.10c), *Myrica* (Abbildung 5.10d) und *Quercus* (Abbildung 5.10e) differenziert werden. Auch hier zeigt sich in den entsprechenden Loadings (Abbildung 5.10f-j), dass dies aus Varianzen gattungsspezifischer Banden (siehe vorigen Abschnitt) mit unterschiedlichen relativen Intensitäten resultiert. Bei der Variation der Banden bei m/z 6814 (Abbildung 5.10g), m/z 4105, 5313, 5930, 6118 und 9045 (Abbildung 5.10h), m/z 5701 und 8969 (Abbildung 5.10i), sowie m/z 6880 (Abbildung 5.10j) handelt es sich jeweils um spezifische Banden der Gattungen *Betula*, *Corylus*, *Myrica* und *Quercus*. Für Spektren der Gattungen *Alnus* und *Quercus* ist die artspezifische Variation dieser Banden so stark

ausgeprägt, dass die Unterscheidung dieser beiden Gattungen mit HCA scheitert (siehe Abbildung 5.6).

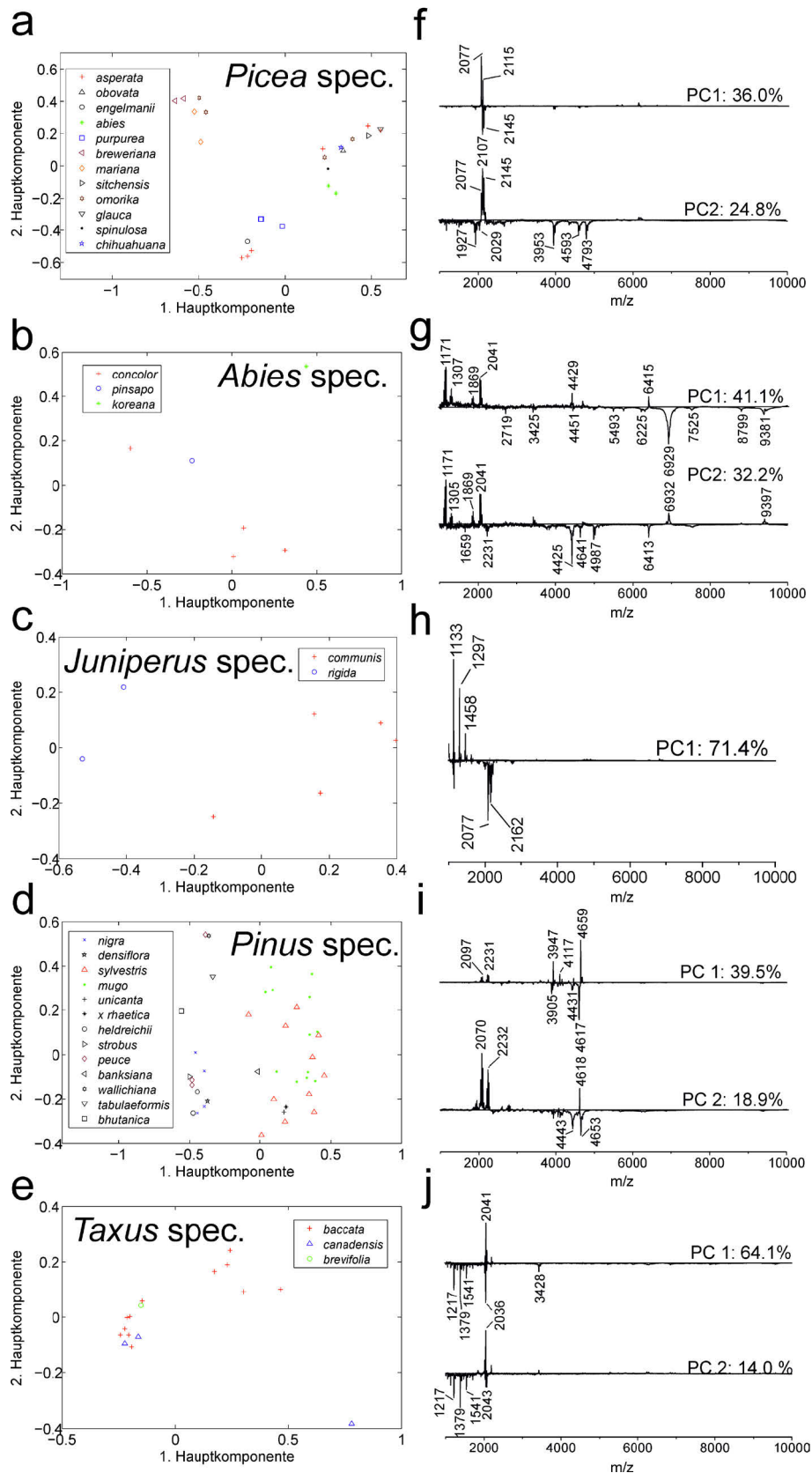


Abbildung 5.9: Ergebnisse der PCA von Pollenspektren verschiedener Gattungen der Ordnung *Coniferales*. Scores der ersten und zweiten Hauptkomponente für *Picea* (a), *Abies* (b), *Juniperus* (c), *Pinus* (d), und *Taxus* (e), sowie die entsprechenden Loadings und Varianzen der jeweiligen Hauptkomponenten (f-j) sind gezeigt.

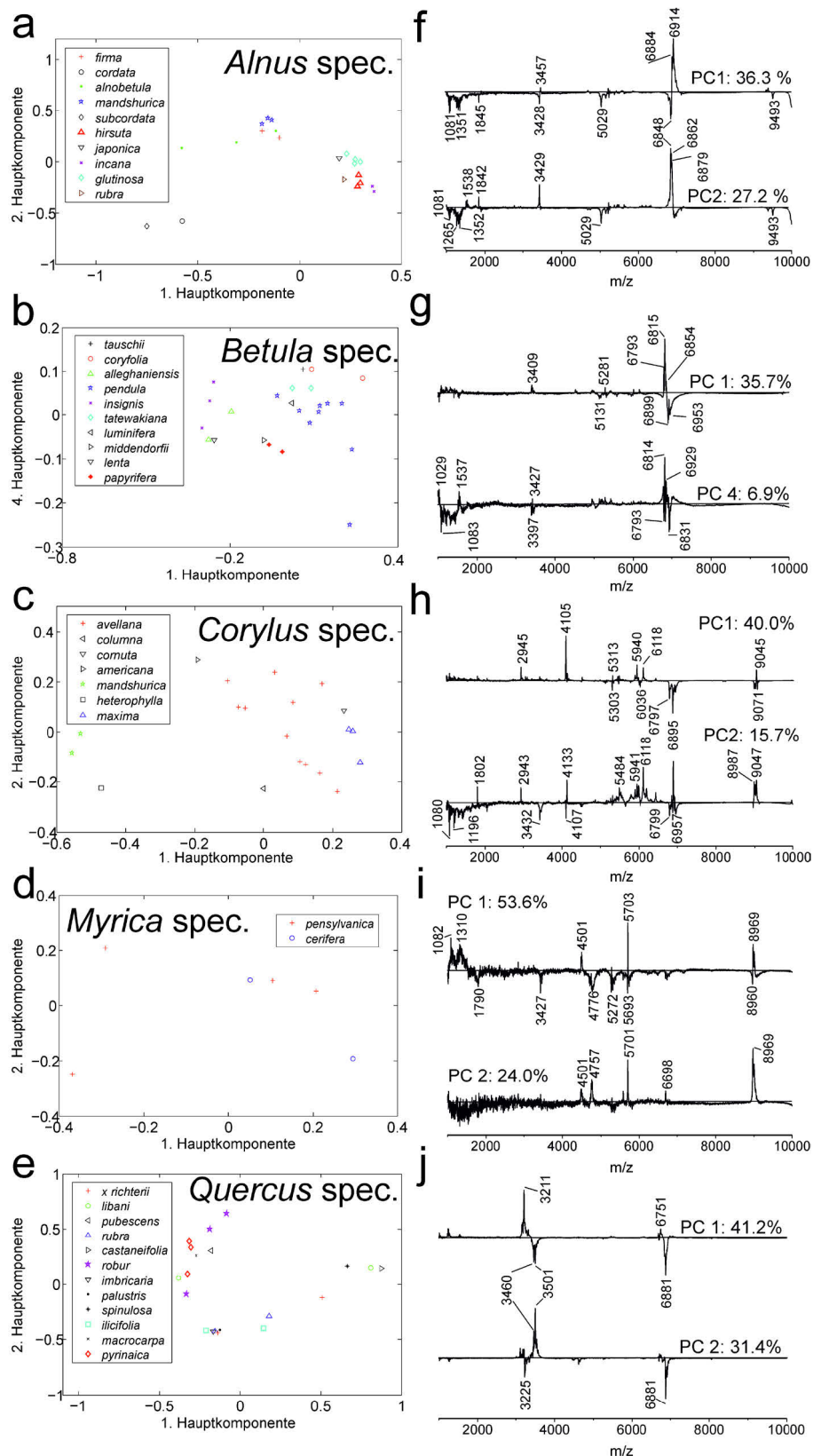


Abbildung 5.10: Ergebnisse der PCA von Pollenspektren verschiedener Gattungen der Ordnung *Fagales*. Scores der ersten und zweiten Hauptkomponente für *Alnus* (a), *Corylus* (c), *Myrica* (d) und *Quercus* (e), Scores für die erste und vierte Hauptkomponente für *Betula* (b), sowie die entsprechenden Loadings und Varianzen der jeweiligen Hauptkomponenten (f-j) sind gezeigt.

5.3 Schlussfolgerungen

Zusammenfassend lässt sich über die multivariate Auswertung von MALDI-TOF MS-Spektren zur taxonomischen Klassifizierung von Pollen sagen, dass der Bereich über m/z 5000 entscheidend für die Unterscheidung der taxonomischen Ordnungen *Fagales* und *Coniferales* ist. Der Massenbereich unter m/z 5000 kann zur gattungsspezifischen Klassifizierung weiter unterteilt werden: Massen unter m/z 2000, die charakteristische Distanzen von 162 Da aufweisen, sind spezifisch für Pollenkörner der Gattung *Juniperus*, während der Massenbereich um m/z 2000 für die Unterscheidung der Gattungen *Taxus*, *Tsuga* und *Picea* verwendet werden kann. Zusätzlich kann zwischen *Pinus* und *Picea*-Pollenspektren anhand spezifischer Massen im Bereich zwischen m/z 3500 und 5000 differenziert werden.

Um die Spektren der Gattungen *Betula* und *Myrica* voneinander, sowie von denen der Gattungen *Quercus*, *Alnus* und *Corylus* zu unterscheiden, sollte der Massenbereich zwischen m/z 6800 und 7000 untersucht werden. Spektren der Gattung *Corylus* weisen weiterhin charakteristische Massen im Bereich zwischen m/z 4000 und 6200 auf, während Spektren von *Quercus* eher charakteristische Banden unterhalb von m/z 4000 besitzen.

In vielen Fällen haben die Massenbereiche und Banden artspezifische Merkmale mit unterschiedlichen relativen Intensitäten, so dass häufig auch zwischen verschiedenen Arten einer Gattung unterschieden werden kann.

Generell kann durch die Kombination von multivariaten Methoden mit MALDI-TOF MS-Daten, äquivalent zur Analyse von Bakterien^{[131],[28]} ein charakteristischer und reproduzierbarer, spektraler Fingerabdruck erhalten werden, um eine taxonomische Klassifizierung von Pollenproben zu erreichen. Ähnlich wie bei der Klassifizierung von Bakterien^[132] und Pilzen^[62] können dabei spektrale Varianzen auf spezifische taxonomische Charakteristika zurückgeführt werden. Diese können anschließend zum Beispiel durch eine PCA systematisch aufgeklärt werden.

6. CPCA von MALDI-TOF MS- und SERS-Daten zur taxonomischen Klassifizierung von Pollen

Bei der *Consensus Principal Component Analysis* (CPCA) handelt es sich um eine multivariate Methode, mit der Daten verschiedener Analyseverfahren für eine umfangreiche Charakterisierung von biologischen Proben kombiniert werden können. Viele aktuelle Publikationen befassen sich mit der Entwicklung und Optimierung der Methode.^{[74],[135]-[137]} Anwendungen auf die Untersuchung von Fleischproben zeigen jedoch das Potential, dass die CPCA für die Untersuchung von biologischen Proben bietet. Es konnten dabei Daten von FTIR-, NIR- und Raman-Spektroskopie kombiniert werden, um detailliert die Wechselwirkung verschiedener Salze auf die Proteine und das Wasser des Fleisches zu untersuchen.^[72] Die Daten dieser spektroskopischen Verfahren konnten anschließend auch mit den Daten anderer, nicht spektroskopisch erfasster Probeneigenschaften der Wasserspeicherkapazität (*water holding capacity*, WHC) und Sensorik ergänzt werden.^[73] In dieser Anwendung wird deutlich, dass eine Kombination von unterschiedlichen, komplementären Informationen durch CPCA sehr hilfreich für das Aufklären komplexer Zusammenhänge sein kann.

Im Rahmen dieser Arbeit soll die CPCA dazu verwendet werden, um Daten der *matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS) und der oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS) zur taxonomischen Klassifizierung zusammenzufügen. Da durch MALDI-TOF MS vor allem Informationen über Glykoproteine und durch SERS vor allem Informationen über Proteine und DNA-Bestandteile erhalten werden, können komplementäre Informationen verknüpft werden, um eine detaillierte Analyse und eine verbesserte Klassifizierung der Pollenproben zu erreichen. Ähnliche Datenverknüpfungen in Multiblock-Analysen spektroskopischer und spektrometrischer Methoden untereinander^{[138],[139]}, sowie miteinander^[140] wurden bereits gezeigt.

Im Folgenden werden die vorbehandelten MALDI-TOF MS-Spektren (siehe Abschnitt 5.2) und die Mittelwertspektren der taxonomisch relevanten SERS-Daten (siehe Abschnitt 4.5) für die CPCA verwendet und die Klassifizierung auf taxonomischer Ebene der Ordnung (Abschnitt 6.1) und Gattung (Abschnitt 6.2) jeweils separat untersucht.

6.1 CPCA zur Klassifizierung auf taxonomischer Ebene der Ordnung

Um eine kombinierte Auswertung von MALDI-TOF MS- und SERS-Daten zur Klassifizierung auf taxonomischer Ebene der Ordnung zu erreichen, wurden die Daten der Methoden mit CPCA ausgewertet. Betrachtet man die Scores der ersten beiden CPCA-Hauptkomponenten (Abbildung 6.1a bis c), so ist erkennbar, dass sowohl durch SERS (Abbildung 6.1a) und MALDI-TOF MS-Block (Abbildung 6.1b), als auch durch den globalen Block (Abbildung 6.1c) zwischen den Proben der beiden Ordnungen differenziert werden kann. Die Loadings der CPCA-Hauptkomponenten (Abbildung 6.1d) sind dabei denen der Hauptkomponentenanalysen der einzelnen Blöcke (siehe Abbildung 4.18b und Abbildung 5.7b) sehr ähnlich. Dies zeigt, dass die ordnungsspezifischen Varianzen der einzelnen Blöcke

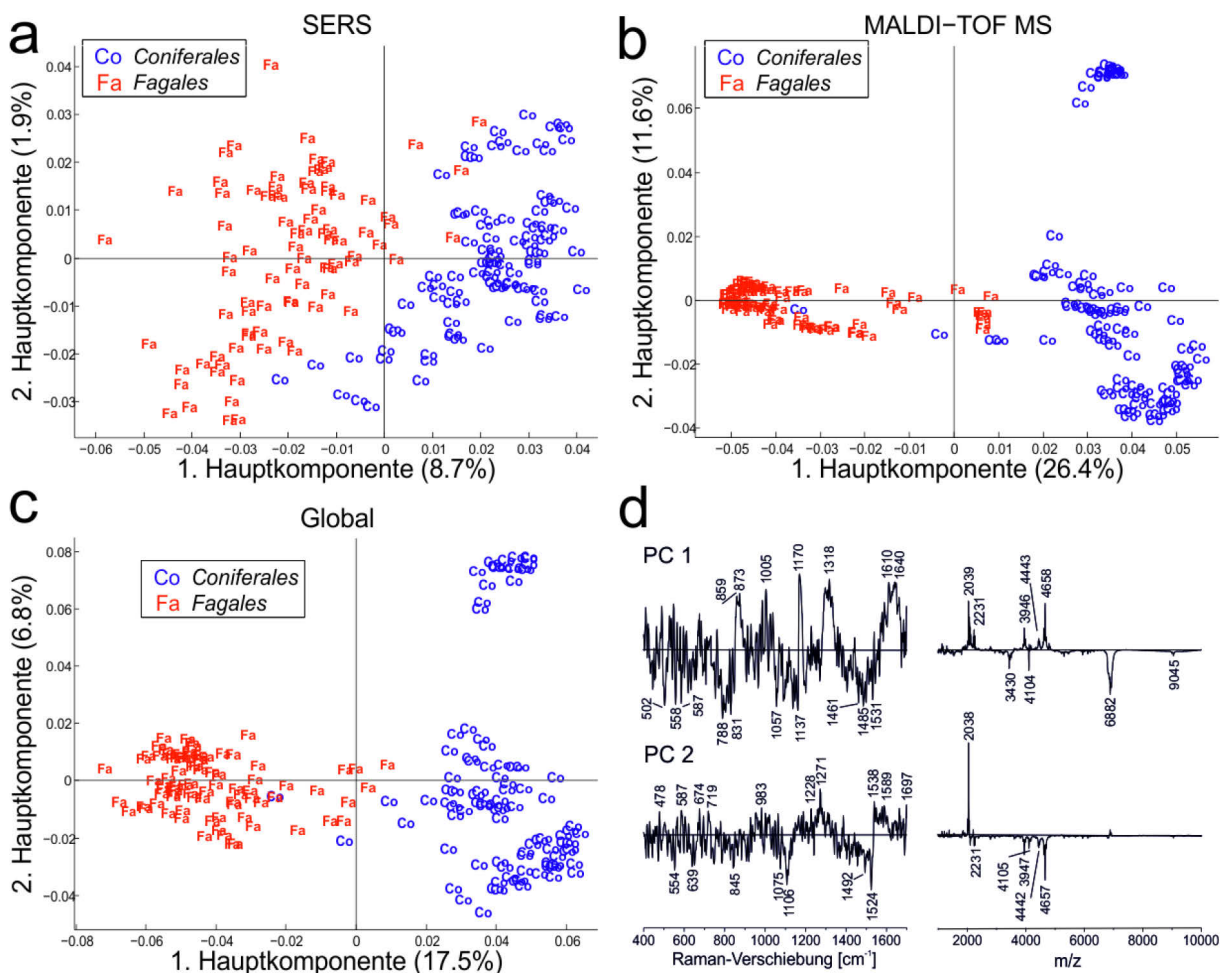


Abbildung 6.1: Ergebnis der CPCA von 180 SERS- und MALDI-TOF MS-Datensätzen zur Klassifizierung auf taxonomischer Ebene der Ordnung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), sowie die Loadings der ersten beiden CPCA-Komponenten (d) sind gezeigt.

in einer CPCA kombiniert werden können. Der Vergleich der Scores des globalen Blocks (Abbildung 6.1c) mit denen des MALDI-TOF MS-Blocks (Abbildung 6.1b), zeigt weiterhin, dass die MALDI-TOF MS-Daten einen großen Einfluss auf die CPCA-Klassifizierung der taxonomischen Ordnung ausüben. Dies wird auch durch die relativ hohen erklärten Varianzen dieses Blocks (Abbildung 6.1b) im Vergleich zu denen des SERS-Blocks (Abbildung 6.1a) bestätigt. Der große Einfluss der MALDI-TOF MS-Daten in der CPCA-Auswertung bewirkt, dass die in der PCA der SERS-Daten (siehe Abschnitt 4.5.3) sichtbaren ordnungsspezifischen Varianzen im SERS-Block noch mehr hervorgehoben werden (Abbildung 6.1a). Anhand der ersten beiden Hauptkomponenten (Abbildung 6.1a) kann also im Vergleich zur PCA des einzelnen Blocks (vgl. Abbildung 4.18) eine klarere Unterscheidung der Datensätze der verschiedenen Ordnungen erfolgen.

Für die erste CPCA-Hauptkomponente ist für den MALDI-TOF MS-Block (Abbildung 6.1b), ähnlich wie für die erste Hauptkomponente der PCA Auswertung (siehe Abbildung 5.7) ein Bereich um den Scoreswert von null vorhanden, bei dem die Klassifizierung der Ordnung einzelner Spektren nicht eindeutig gegeben ist. Betrachtet man die Scores der ersten beiden Hauptkomponenten des globalen Blocks (Abbildung 6.1c), ist erkennbar, dass anhand dieser die Differenzierung in diesem Bereich eindeutiger ist. Selbst der vergleichsweise geringere Einfluss des SERS-Blocks bewirkt in der CPCA-Auswertung also, dass die Klassifizierung auf der taxonomischen Ebene der Ordnung robuster wird und damit der Klassifizierungserfolg verglichen mit dem der PCA der MALDI-TOF MS-Daten (vgl. Abbildung 5.7) gesteigert werden kann.

Der Einfluss des MALDI-TOF MS-Blocks bewirkt, dass ordnungsspezifische Varianzen aus den SERS-Daten herausgearbeitet werden können und damit der Einfluss anderer, z.B. präparationsbedingter Varianzen verringert wird. Somit kann in diesem Fall das unüberwachte Verfahren CPCA dazu verwendet werden, um einen Datenblock mit Hilfe eines anderen zu überwachen und die für die taxonomische Klassifizierung relevanten Informationen aus dem Datensatz zu extrahieren. In Abschnitt 4.5 wurde gezeigt, dass für die Extraktion taxonomisch relevanter Spektren neuronale Netze verwendet werden können. Die Mittelwertspektren dieser selektierten Daten wurden wie in Abbildung 6.1 gezeigt in einer CPCA mit MALDI-TOF MS-Daten verknüpft. Als nächster Schritt soll nun überprüft werden, ob die Extraktion taxonomisch relevanter Varianzen auch dann durch die CPCA erfolgreich durchgeführt werden kann, wenn SERS-Daten verwendet werden, die nicht zuvor durch ein ANN bewertet und selektiert wurden. Abbildung 6.2 zeigt die Ergebnisse der CPCA dieser Daten. Auch anhand der Scores dieses SERS-Blocks (Abbildung 6.2a) ist eine grobe Unterscheidung

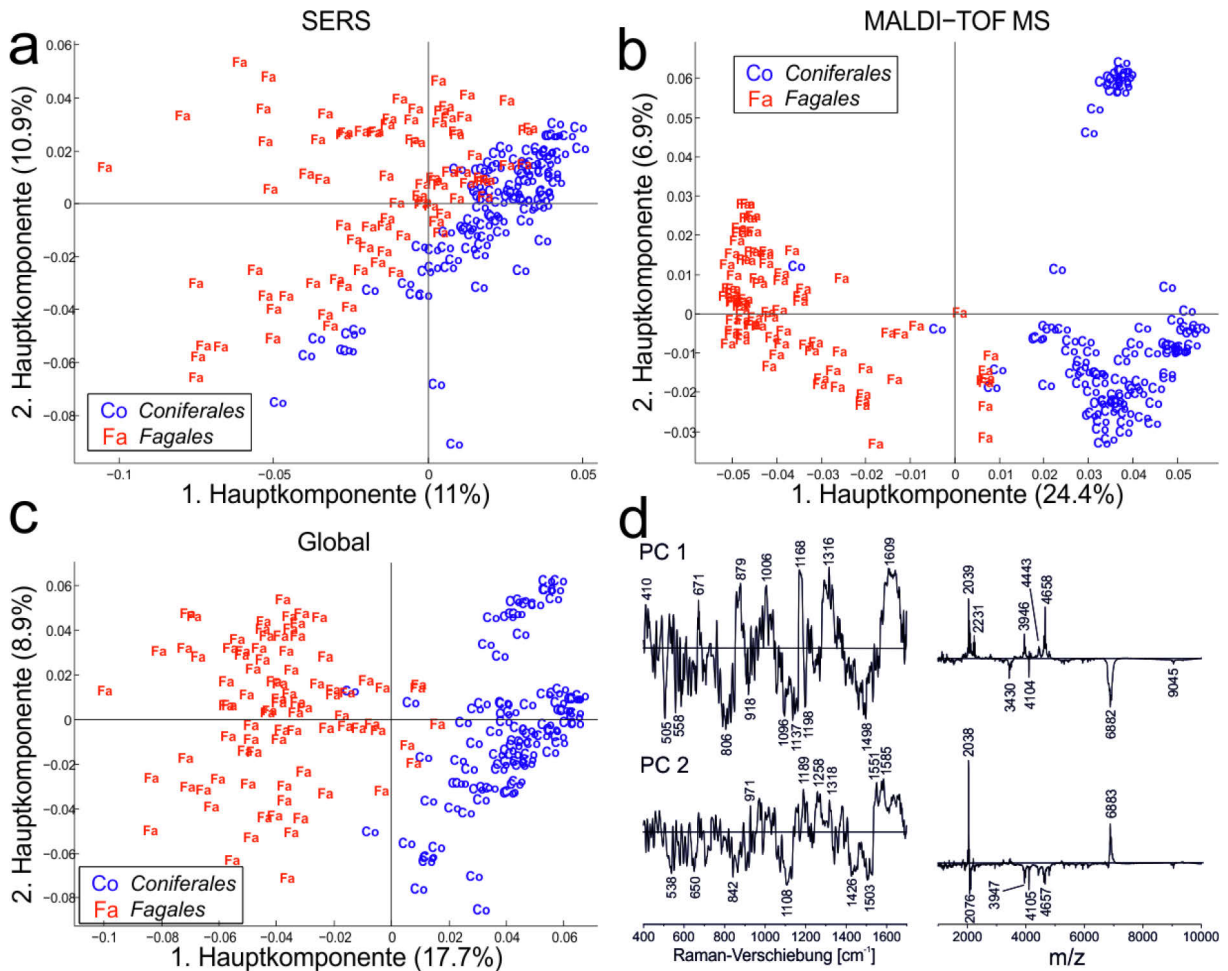


Abbildung 6.2: Ergebnis der CPCA von 180 Datensätzen nicht selektierter SERS- und MALDI-TOF MS-Spektren zur Klassifizierung auf taxonomischer Ebene der Ordnung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), sowie die Loadings der ersten beiden CPCA-Komponenten (d) sind gezeigt.

zwischen den Datensätzen der verschiedenen Ordnungen möglich, und taxonomisch relevante Varianzen können somit zumindest teilweise auch in dieser Auswertung beobachtet werden. Verglichen mit der CPCA der extrahierten Daten (vgl. Abbildung 6.1) ist diese Unterscheidung jedoch sowohl in den Scores der einzelnen Blöcke (Abbildung 6.2a und b), als auch in denen des globalen Blocks (Abbildung 6.2c) weniger eindeutig. Dies ist darauf zurückzuführen, dass der SERS-Block für beide CPCA-Hauptkomponenten einen größeren Einfluss ausübt. Er weist nämlich größere erklärte Varianzen für die CPCA-Hauptkomponenten auf (vgl. Abbildung 6.2a mit Abbildung 6.1a). Damit werden taxonomisch irrelevante Varianzen der SERS-Daten stärker in die Auswertung mit einbezogen, was generell die taxonomische Klassifizierung beeinträchtigt. Dies ist auch an den signifikant unterschiedlichen Loadings des SERS-Blocks der beiden Hauptkomponenten

ersichtlich (vgl. Abbildung 6.2d mit Abbildung 6.1d). Dass trotzdem sowohl durch die globalen Scores (Abbildung 6.2c) als auch durch die Scores des SERS-Blocks (Abbildung 6.2a) viele Datensätze verschiedener Ordnungen differenziert werden können, zeigt, dass sogar nicht selektierte SERS-Daten für die Klassifizierung mit CPCA verwendet werden können.

6.2 CPCA zur Klassifizierung auf taxonomischer Ebene der Gattung

Für die CPCA-Klassifizierung auf taxonomischer Ebene der Gattung wurden analog zur PCA der einzelnen Blöcke (siehe Abschnitt 4.5.5 und Abschnitt 5.2.2) die Daten der beiden Ordnungen einzeln untersucht. Bei der CPCA der Datensätze der Ordnung *Coniferales* können dabei die globalen Scores der ersten vier CPCA-Hauptkomponenten für eine Klassifizierung der verschiedenen Gattungen verwendet werden (siehe Abbildung 6.3c und f). Dabei sind die positiven Scores der ersten CPCA-Hauptkomponente maßgeblich, um die Daten von *Taxus* von denen der anderen Gattungen zu unterscheiden (siehe Abbildung 6.3c). In den positiven Loadings dieser Hauptkomponente (Abbildung 6.3g, erste Spur) können die für diese Gattung charakteristische MALDI-TOF MS-Peaks bei m/z 2039 und 2201 (vgl. Abbildung 5.3) identifiziert werden. Der Vergleich der globalen Scores (Abbildung 6.3c) dieser CPCA-Hauptkomponente mit der des reinen MALDI-TOF MS-Blocks (Abbildung 6.3b), sowie die hohe erklärte Varianz von über 25% für diesen Block (Abbildung 6.3b) verglichen mit nur 5.8% für den SERS-Block (Abbildung 6.3a) zeigt, dass diese Komponente größtenteils durch den MALDI-TOF MS-Block beeinflusst wird. Dies führt dazu, dass *Taxus*-spezifische Varianzen auch im SERS-Block intensiv gewichtet werden, und auch eine Identifizierung vieler *Taxus*-Datensätze anhand der SERS-Daten erfolgen kann (Abbildung 6.3a). Bei der PCA dieses Blocks war dies nur ansatzweise möglich (vgl. Abbildung 4.22a bis c). Ähnlich wie die erste ist auch die zweite CPCA-Hauptkomponente, die zur Unterscheidung von positiven Scores der Gattung *Pinus* von negativen Scores der Gattung *Picea* verwendet werden kann (Abbildung 6.3c), primär von den Daten des MALDI-TOF MS-Blocks beeinflusst. Dies kann auch bei dieser Komponente aus ähnlichen Scores dieses Blocks (Abbildung 6.3b), verglichen mit denen des globalen Blocks (Abbildung 6.3c), sowie anhand der vergleichsweise hohen erklärten Varianz von fast 20 % (Abbildung 6.3b) gegenüber nur 5.2% des SERS-Blocks (Abbildung 6.3a) geschlossen werden. Die Loadings (Abbildung 6.3g, zweite Spur) zeigen charakteristische MALDI-TOF MS-Peaks, z.B. Maxima bei m/z 3945 und 4657 und Minima bei m/z 2076 und 3955, welche typischen Peaks von *Pinus*-, bzw. *Picea*-Pollenproben zugeordnet werden können (vgl. Abbildung 5.3). Auch bei der Unter-

scheidung von Datensätzen der Gattungen *Pinus* und *Picea* anhand der zweiten CPCA-Hauptkomponente (Abbildung 6.3a) können gattungsspezifische Varianzen in den SERS-Daten hervorgehoben werden. Dies unterscheidet die CPCA-Auswertung der SERS-Daten von der PCA des reinen Blocks (vgl. Abbildung 4.22a bis c), da dort keine Varianzen identifiziert werden konnten, die die Unterscheidung dieser beiden Gattungen ermöglichten.

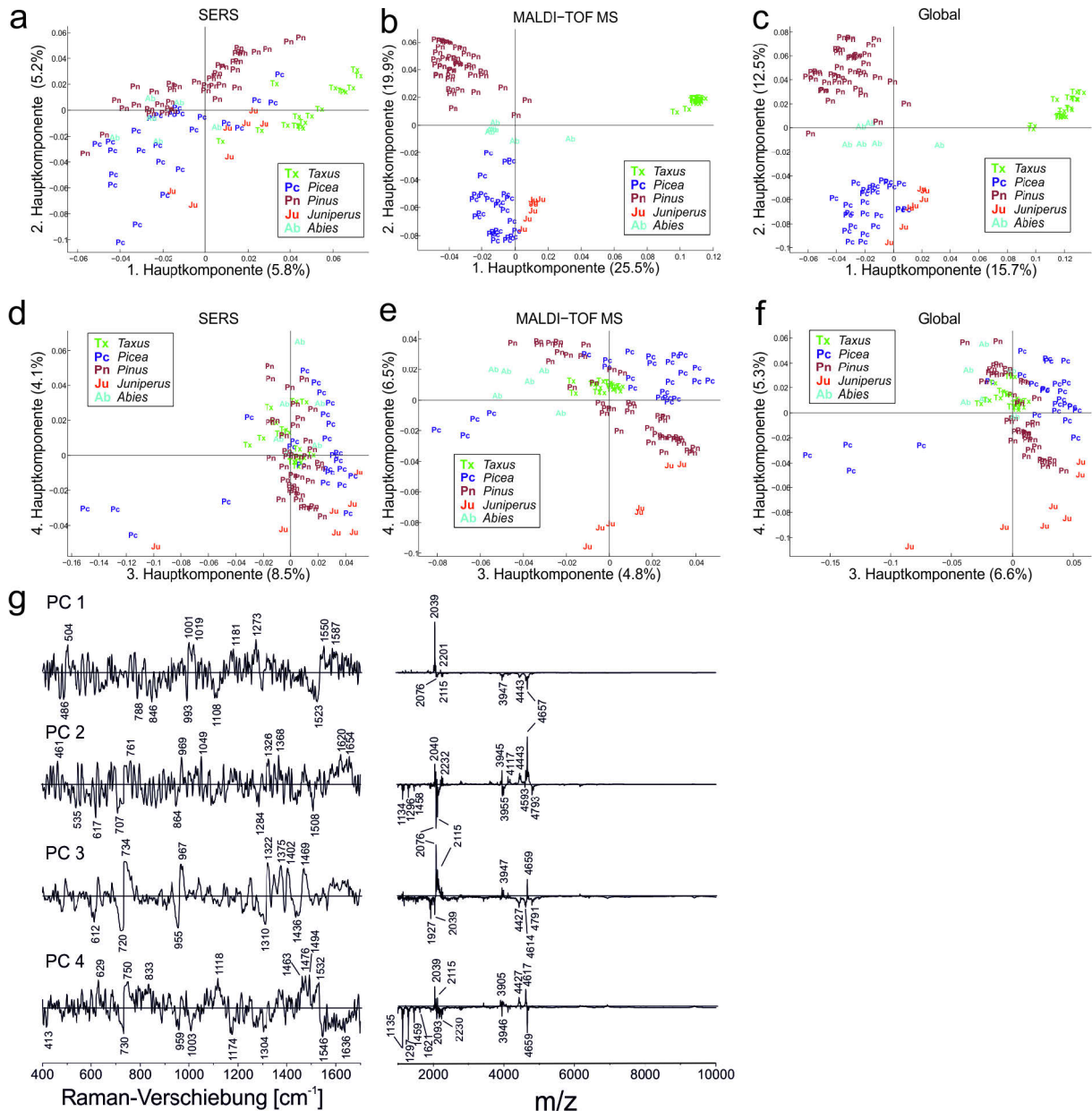


Abbildung 6.3: Ergebnisse der CPCA von 95 SERS- und MALDI-TOF MS-Datensätzen der Ordnung *Coniferales* zur Klassifizierung auf taxonomischer Ebene der Gattung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponenten für die einzelnen Blöcke (a+b) und den globalen Block (c), die Scores und Varianzen der dritten und vierten CPCA-Komponenten für die einzelnen Blöcke (d+e) und den globalen Block (f), sowie die Loadings der jeweiligen CPCA-Komponenten (g) sind gezeigt.

Für die Datensätze der *Abies*-Pollenproben sind Scores von ca. null für die ersten beiden CPCA-Hauptkomponenten charakteristisch (Abbildung 6.3c). In den SERS-Daten dieser Hauptkomponente (Abbildung 6.3a) sind allerdings keine spezifischen Scores für diese Gattung detektierbar. Charakteristische Merkmale von *Abies*-Pollenproben können also selbst durch Kombination mit MALDI-TOF MS-Daten nicht aus den SERS-Daten hervorgehoben werden. Die Datensätze dieser Gattung sind demnach denen der anderen *Pinaceae*-Pollenproben sehr ähnlich. Dies wird auch in der mangelnden Klassifizierung dieser Gattung mit ANN (siehe Abbildung 4.11) deutlich.

Um Pollenproben der Gattung *Juniperus* von denen der anderen Gattungen zu unterscheiden, sind die negativen Scores der vierten CPCA-Hauptkomponente entscheidend (Abbildung 6.3f). Dies ist auch daran ersichtlich, dass charakteristische Peaks dieser Gattung z.B. bei m/z 1135, 1297 oder 1459 in den Loadings des MALDI-TOF MS-Blocks (Abbildung 6.3g, vierte Spur) auftreten. Die ähnlichen Werte der erklärten Varianz von 4.1% für den SERS-Block (Abbildung 6.3d), bzw. 6.5% für den MALDI-TOF MS-Block (Abbildung 6.3e), sowie die Scores der CPCA-Komponente (Abbildung 6.3d und e) zeigen, dass beide Datenblöcke zu dieser Differenzierung beitragen.

Die dritte CPCA-Hauptkomponente ist die einzige, auf die die SERS-Daten mit einer erklärten Varianz von 8.5% (Abbildung 6.3d) einen größeren Einfluss ausüben als die MALDI-TOF MS-Daten, die nur durch eine erklärte Varianz von 4.8% in dieser CPCA-Hauptkomponente repräsentiert sind (Abbildung 6.3e). Betrachtet man die Loadings dieser Hauptkomponente (Abbildung 6.3g, dritte Spur) so sind charakteristische SERS-Banden von Adenin und Adeninderivaten erkennbar. Die dritte CPCA-Hauptkomponente repräsentiert demnach v.a. Varianzen, die nicht gattungsspezifisch sind und bereits bei der Analyse des SERS-Blocks mit PCA (vgl. Abbildung 4.22) als solche zugeordnet wurden. Da durch die globalen Scores der anderen CPCA-Hauptkomponenten bereits alle Gattungen differenziert werden konnten, kann diese Komponente bei der Klassifizierung vernachlässigt werden.

Auch anhand der CPCA-Analyse der Datensätze der Ordnung *Fagales* ist eine Klassifizierung auf taxonomischer Ebene der Gattung möglich (Abbildung 6.4). Die erste CPCA-Hauptkomponente separiert dabei negative Scores von Datensätzen der Gattungen *Alnus* und *Quercus* von positiven Scores der Gattungen *Corylus* und *Myrica* (Abbildung 6.4c). Sie wird dabei von beiden Datenblöcken beeinflusst, was anhand von jeweils relativ hohen erklärten Varianzen von 15% (Abbildung 6.4a) bzw. 20.1% (Abbildung 6.4b) der jeweiligen Blöcke abgelesen werden kann. Deshalb ist diese Differenzierung auch anhand der Scores der einzelnen Blöcke möglich. In den Loadings der MALDI-TOF MS-Daten (Abbildung 6.4g,

erste Spur) können jeweils die gattungsspezifischen Banden bei m/z 4105 und m/z 5701 für Proben der Gattung *Corylus* und Myrica, und um m/z 6880 für solche der Gattung *Alnus* und *Quercus* identifiziert werden.

Die zweite CPCA-Hauptkomponente (Abbildung 6.4c) ist durch Unterscheidung negativer Scores von Datensätzen der Gattung *Betula* von denen positiverer Werte anderer Gattungen

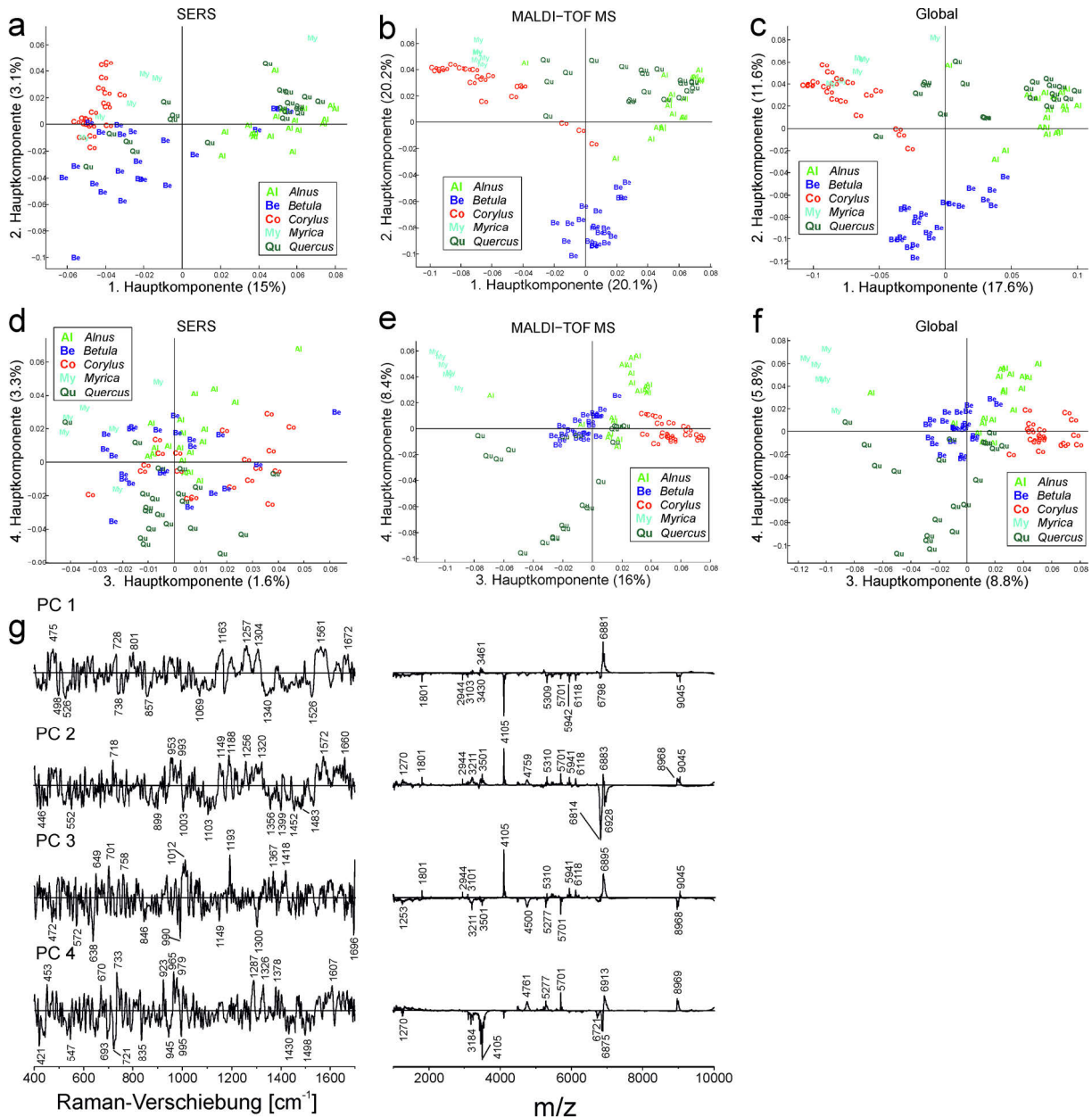


Abbildung 6.4: Ergebnisse der CPCA von 85 SERS- und MALDI-TOF MS-Datensätzen der Ordnung *Fagales* zur Klassifizierung auf Ebene der taxonomischen Gattung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), die Scores und Varianzen der dritten und vierten CPCA-Komponente für die einzelnen Blöcke (d+e) und den globalen Block (f), sowie die Loadings der jeweiligen CPCA-Komponenten (g) sind gezeigt.

der Ordnung *Fagales* charakterisiert. Deshalb können auch typische Peaks von *Betula*-Pollenproben, wie zum Beispiel bei m/z 6815 in den negativen Loadings der MALDI-TOF MS-Daten (Abbildung 6.4g, zweite Spur) identifiziert werden. Die Ähnlichkeiten der Scores des globalen Blocks (Abbildung 6.4c) und des MALDI-TOF MS-Blocks (Abbildung 6.4b), sowie die hohe erklärte Varianz dieses Blocks (Abbildung 6.4b) verglichen mit der des SERS-Blocks (Abbildung 6.4a) zeigt, dass diese Differenzierung primär durch MALDI-TOF MS-Daten beeinflusst ist. Die Scores des SERS-Blocks (Abbildung 6.4a) zeigen jedoch, dass auch typische Varianzen von Spektren der Gattung *Betula* aus den SERS-Daten herausgearbeitet werden können. Eine Differenzierung der *Betula*-Pollenproben von denen anderer Gattungen war anhand der PCA des reinen SERS-Blocks (vgl. Abbildung 4.22) nicht möglich. Auch hier führt also der Einfluss des MALDI-TOF MS-Blocks dazu, dass gattungsspezifische Varianzen im SERS-Block stärker gewichtet werden können.

Ähnliches kann auch bei der dritten CPCA-Hauptkomponente (Abbildung 6.4f), die zwischen positiven Scores der *Corylus*-Datensätze und negativen Scores der *Myrica*-Datensätze differenziert, beobachtet werden. Auch diese CPCA-Hauptkomponente ist primär durch den MALDI-TOF MS-Block (Abbildung 6.4e) beeinflusst und charakteristische Peaks von *Corylus*-Pollenproben bei m/z 4105 und 9045 können in den positiven Loadings (Abbildung 6.4g, dritte Spur), sowie typische Peaks von *Myrica*-Pollenproben bei m/z 4105 und 9045 in den negativen Loadings (Abbildung 6.4g, dritte Spur) dieser CPCA-Hauptkomponente wiedergefunden werden. Der große Einfluss dieses Blocks ist sowohl an der bedeutend größeren erklärten Varianz von 16% (Abbildung 6.4e) verglichen mit nur 1.6% des SERS-Blocks (Abbildung 6.4d), als auch an der großen Ähnlichkeit der Scores dieses Blocks (Abbildung 6.4e) mit denen des globalen Blocks (Abbildung 6.4f) ersichtlich. Eine Differenzierung zwischen Datensätzen von *Corylus*- und *Myrica*-Pollenproben war anhand des reinen SERS-Blocks (vgl. Abbildung 4.22) nicht möglich, kann aber nun bei der CPCA-Auswertung durch den Einfluss der MALDI-TOF MS-Daten in der dritten CPCA-Hauptkomponente erreicht werden (Abbildung 6.4d).

Anhand der vierten CPCA-Hauptkomponente (Abbildung 6.4f) ist die Unterscheidung zwischen positiven Scores von Datensätzen der Gattung *Quercus* und negativen Scores von Datensätzen der Gattung *Alnus* möglich. Die gattungsspezifischen Varianzen können in der CPCA-Auswertung also verknüpft werden und finden sich auch in den Scores der einzelnen Blöcke der CPCA (Abbildung 6.4d und e) wieder. Vergleicht man die Scores der einzelnen Blöcke (Abbildung 6.4d und e) mit denen des globalen Blocks (Abbildung 6.4f), ist erkennbar, dass die Kombination der Daten zu robusteren Klassifizierungsergebnissen führt.

Dies wird besonders durch die Unterscheidung um einen Scoreswert von null illustriert (Abbildung 6.4f).

6.3 Schlussfolgerungen

Durch die CPCA-Untersuchung von Datensätzen verschiedener taxonomischer Ordnungen und Gattungen konnte gezeigt werden, dass SERS- und MALDI-TOF MS-Daten für eine robustere Klassifizierung erfolgreich verknüpft werden können. Dies beweist, dass auch diese Methoden, ähnlich wie FTIR-, NIR- und normale Raman-Spektroskopie^[72] für eine umfassende Charakterisierung und Klassifizierung komplexer, biologischer Proben kombiniert werden können. Der hohe Einfluss der MALDI-TOF MS-Daten führt bei der CPCA dazu, dass taxonomisch relevante Varianzen in den SERS-Daten stärker gewichtet und damit herausgearbeitet werden können. Während die taxonomische Klassifizierung der SERS-Daten nur teilweise anhand der PCA erreicht werden konnte (siehe Abschnitt 4.5), können nun experiment- und probenspezifische Varianzen, die gattungsspezifischen Varianzen überlagern, durch diese Kombination erfolgreich unterdrückt werden. Dies zeigt, dass sich unüberwachte CPCA nicht nur zum Kombinieren der Daten eignet, sondern auch, dass einzelne Datenblöcke dazu verwendet werden können, um andere zu überwachen, und die für die Klassifizierung relevanten Varianzen sichtbar zu machen. Im Gegensatz zu der direkten Überwachung, die häufig bei Multiblock-Auswertung angewendet wird^{[140],[138]}, werden dabei allerdings keine direkten Klasseninformationen in die Auswertung mit CPCA eingebracht.

7. Klassifizierung von FTIR-Daten einzelner Pollenkörner

Bei der Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie) handelt es sich um eine spektroskopische Methode, die bereits in zahlreichen Anwendungen für die Untersuchung von biologischen Proben benutzt wurde. Dabei können die FTIR-Spektren als spektrale Fingerabdrücke und damit, kombiniert mit multivariaten Verfahren, zur Charakterisierung der Proben verwendet werden. Diese leistungsfähige Kombination wird zum Beispiel für die Klassifizierung von Lebensmitteln, wie Honig^{[141],[142]}, Fruchtsaft^[143], Rotwein^[144], Milch^[145] oder Öl^{[146],[147]} angewendet. Da sich die FTIR-Spektroskopie besonders für die Untersuchung von Proteinen eignet,^[148] hat sich diese Kombination auch in der medizinischen Forschung etabliert.^{[33],[34]} So können beispielsweise verschiedene Typen von Kollagen differenziert^[149] und Krebszellen von gesunden Zellen unterschieden werden^{[150],[151]}. Besonders bei letzterer Anwendung zeigt sich, dass auch chemische Karten von Gewebeproben mit dieser kombinierten Methode erhalten werden können.^{[152],[153]} Auch pflanzliche Proben, wie z.B. Algen^[154] oder Holz^[155] wurden mit FTIR-Spektroskopie und multivariaten Methoden bereits klassifiziert. Das Potential für die taxonomische Klassifizierung von biologischen Proben wurde ebenfalls eindrücklich, u.a. anhand der Untersuchung von Mikroorganismen^{[35],[156]} oder Blättern^[157] gezeigt.

Bei der Anwendung auf Pollenproben wurde deutlich, dass diese sowohl von Sporen unterschieden werden können,^[26] als auch dass eine Klassifizierung verschiedener taxonomischer Gruppen möglich ist^[16]. In weiteren Experimenten zeigte sich, dass sogar Umwelteinflüsse, die auf die Pflanze wirken, anhand von Untersuchungen der Pollen detektiert werden können.^[17] Um diese Methoden für die Klassifizierung von Pollen zu etablieren, sollten allerdings im Gegensatz zu den bisherigen Anwendungen, bei denen Kaliumbromidpresslinge verwendet oder die Messungen unter abgeschwächter Totalreflexion (*attenuated total reflection, ATR*) durchgeführt wurden, einzelne Pollenkörner durch Transmissionsmessungen am Mikroskop analysiert werden. Die Herausforderung bei einem derartigen Versuchsaufbau ist jedoch, dass Partikel untersucht werden, die eine Größe von 10 bis 100 μm aufweisen und damit im Bereich der Wellenlänge des Infrarotlichtes liegen. Ist dies der Fall, so kommt es bei FTIR-Messungen zur Überlagerung der Absorptionsbeiträge mit Beiträgen, die auf Streuung an den Partikeln zurückgeführt werden können und damit von der Morphologie, Größe und optischer Dichte der Partikel abhängen.^[41] Im Spektrum äußern sich diese Beiträge zum einen durch einen breiten, oszillierenden Untergrund, der aus Mie-Streuung an den Partikeln resultiert^[158] und vor allem durch die Größe der Partikel beeinflusst

wird.^[41] Zum anderen treten auch sogenannte Dispersionsartefakte auf, die die Veränderung der Form einzelner Absorptionsbanden bewirken und dadurch verursacht werden, dass die Streuung auch durch eine Änderung des Brechungsindex beeinflusst wird.^[41] Um aus den Spektren die Absorptionsbeiträge zu extrahieren, wurde aus dem EMSC (*Extended Multiplicative Signal Correction*) Algorithmus, der zur allgemeinen Streukorrektur von FTIR-Spektren entwickelt wurde^[39] (siehe Abschnitt 2.1.1) ein Algorithmus abgeleitet, um den Einfluss der resonanten Mie-Streuung, also auch den von Dispersionsartefakten im Spektren zu reduzieren.^{[42]-[45],[159]} Erste Anwendungen an Gewebeproben zeigen, dass tatsächlich eine Verringerung des Einflusses der resonanten Mie-Streuung auf die Spektren erreicht werden kann^{[160],[161]}, und anhand von korrigierten FTIR-Spektren z.B. auch Zellen in verschiedenen Zellzyklen differenziert werden können^[162].

Das Ziel dieser Arbeit ist es, reproduzierbare und spezifische FTIR-Daten einzelner Pollenkörner zu erhalten und diese anschließend für die taxonomische Klassifizierung zu verwenden. Eine Extraktion der reinen Absorptionsbeiträge aus den FTIR-Spektren von Pollenkörnern ist allerdings aufgrund der hochkomplexen Streuartefakte, die diese überlagern, aktuell noch nicht möglich.^[27] Ein Algorithmus, durch den sich der Einfluss von breiten, oszillierenden Banden, die durch Streuung verursacht werden, und schmalen, hauptsächlich durch Absorption beeinflussten Banden in der Datenauswertung variieren lässt, ist, wie von Zimmermann et al. gezeigt,^[163] die systematische Änderung der Anzahl an Savitzky-Golay (SG) Datenpunkten bei der Bildung der zweiten Ableitung kombiniert mit einer EMSC-Korrektur der Spektren. Durch Anwendung dieses Algorithmus‘ auf die Daten einzelner Pollenkörner kann dabei nicht nur der Einfluss der oszillierenden Untergrundbanden, sondern auch der Einfluss einzelner Absorptionsbanden, die unterschiedliche Peakbreiten aufweisen, variiert und für die Klassifizierung optimiert werden. Zur Klassifizierung der FTIR-Spektren sollen hierarchische PLSR-basierte Klassifizierungsmodelle^[164] aufgebaut werden, an dessen Knotenpunkten jeweils die optimale Anzahl an SG-Datenpunkten bestimmt wird. Ähnlich wie bei den Untersuchungen mit MALDI-TOF MS und SERS sollen dafür zunächst kommerziell erhältliche Pollenproben verschiedener Arten untersucht werden, um zu überprüfen, ob reproduzierbare, spezifische Pollendaten erhalten werden können (Abschnitt 7.1). Anschließend sollen im Feld gesammelte Proben analysiert werden, um zu validieren, ob taxonomische Verwandtschaftsverhältnisse anhand der FTIR-Daten reproduziert werden können (Abschnitt 7.2).

7.1 Eignung der FTIR-Spektroskopie für die Klassifizierung einzelner Pollenkörner

Um zu überprüfen, ob auch FTIR-Spektren einzelner Pollenkörner zur taxonomischen Klassifizierung von Pollenproben verwendet werden können, muss zunächst untersucht werden, ob FTIR-Daten erhalten werden, die reproduzierbaren taxonomisch spezifischen Fingerabdrücken der Pollenproben entsprechen. Zu diesem Zweck wurden kommerziell erhältliche Pollenproben (siehe Tabelle 3.1) verschiedener Arten untersucht. Abbildung 7.1 zeigt repräsentative FTIR-Spektren von einzelnen Pollenkörnern der Arten *Juniperus scopulorum* (a), *Populus nigra* (b) und *Secale cereale* (c). Es können Absorptionsbanden identifiziert werden, die auch bei den Untersuchungen von Kaliumbromidpresslingen^{[16],[17]} auftreten (Tabelle 7.1). Diese können unterschiedlich breit sein. Die Carbonylstreckschwingung von Lipiden bei ca. 1740 cm^{-1} ist beispielsweise relativ schmal, während z.B. die Amid I- und Amid II-Banden, aufgrund der Überlagerung mehrerer Bandenkomponenten, relativ breit erscheinen. Es ist allerdings auch erkennbar, dass sich Streuartefakte in den Spektren befinden. So weisen einzelne Absorptionsbanden teilweise sehr

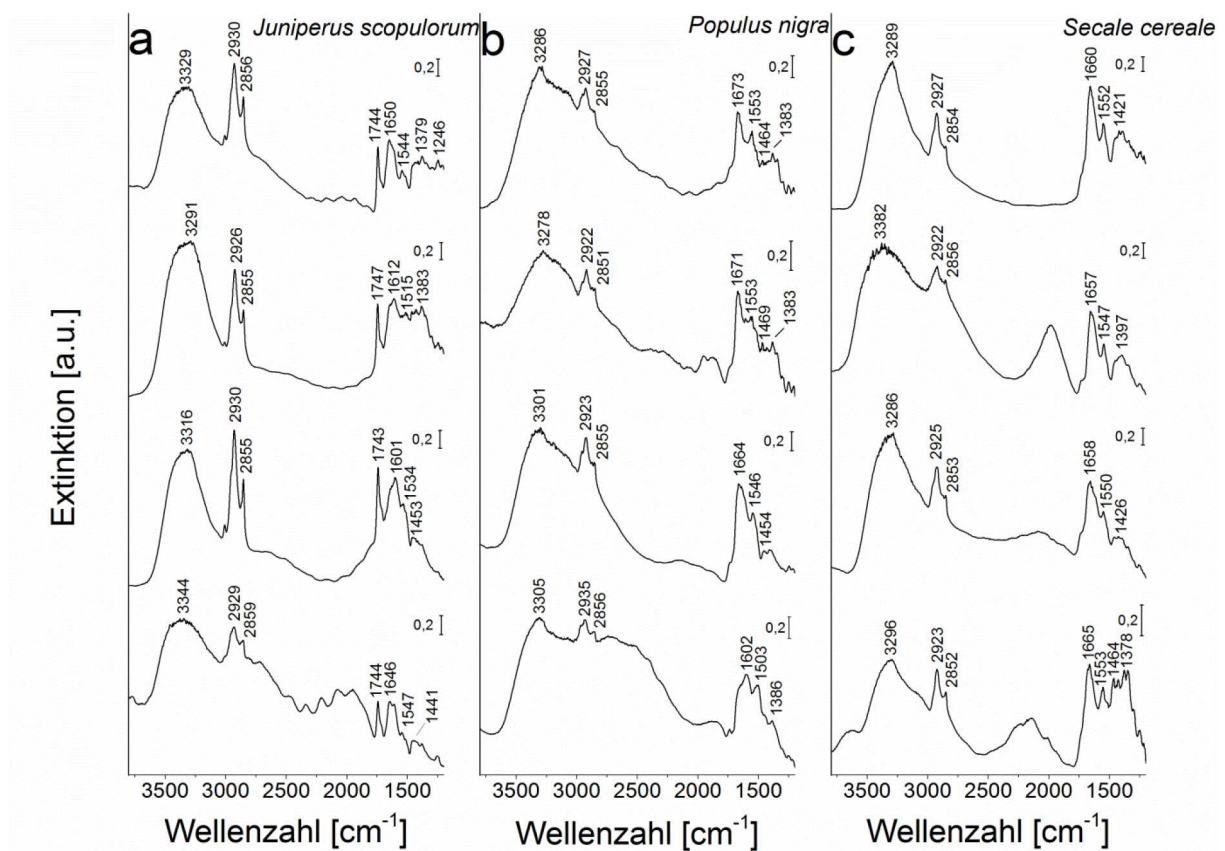


Abbildung 7.1: Repräsentative FTIR-Spektren einzelner Pollenkörner der Arten *Juniperus scopulorum* (a), *Populus nigra* (b) und *Secale cereale* (c). Für eine Zuordnung der Absorptionsbanden siehe Tabelle 7.1.

Tabelle 7.1: Charakteristische Absorptionsbanden der FTIR-Spektren einzelner Pollenkörner und deren Zuordnung zu Referenzspektren

Wellenzahl [cm^{-1}]	Zuordnung ^a
3266-3382	ν , OH
2922-2935	ν , CH_3 asym.
2851-2859	ν , CH_3 sym.
1743-1747	ν , C=O, Lipide ^b
1601-1673	ν , C=O, Proteine, Amid I ^b
1503-1553	ν , C-N, δ , N-H, Proteine, Amid II ^b
1441-1469	δ , CH_2 , Lipide ^b
1378-1386	δ , OH
1246	ν , C-O, Lipide ^b

^a Abkürzungen: ν , Streckschwingung δ , Deformationsschwingung
sym, symmetrisch asym, asymmetrisch

^b nach [16] und [17]

stark unterschiedliche Maxima auf, was zum Beispiel an der Amid I-Bande der Spektren der Arten *Juniperus scopulorum* (Abbildung 7.1a) und *Populus nigra* (Abbildung 7.1b) ersichtlich ist. Dies ist vermutlich sowohl auf oszillierende Untergrundbanden, als auch auf Dispersionsartefakte^[41] zurückzuführen. Zusätzlich sind Banden in spektralen Regionen sichtbar, in denen keine Absorptionsbanden auftreten. Dies gilt zum Beispiel für den Bereich um 2000 cm^{-1} (siehe z.B. Abbildung 7.1a, vierte Spur oder Abbildung 7.1c, zweite Spur) bei diesen handelt es sich um oszillierende Untergrundbanden, die durch Mie-Streuung an den Pollenkörnern verursacht werden. Vergleicht man die Spektren der jeweils gleichen Pollenart, so ist erkennbar, dass zwar spezifische spektrale Merkmale auftreten, aber auch, dass durch eine Auswertung der Daten ohne weitere Bearbeitung keine spezifische Klassifizierung erfolgen kann. Deshalb wurde nach Vorbild von Liland *et. al.*^[164] ein hierarchisches Klassifizierungsmodell aufgebaut, das eine Optimierung der SG-Datenpunkten bei der Spektrenvorbereitung nach Zimmermann *et. al.*^[163] an den jeweiligen Knotenpunkten beinhalten soll. Durch die Variation der Anzahl der SG-Datenpunkte bei der zweiten Ableitung soll dabei nicht nur der Einfluss von breiten, oszillierenden Untergrundbanden, sondern auch der von Absorptionsbanden, die eine unterschiedliche Breite aufweisen, variiert werden. Da hierbei zunächst nicht primär taxonomische Verwandtschaftsverhältnisse abgebildet, sondern generell überprüft werden soll, ob spezifische, reproduzierbare Fingerabdrücke der Pollenproben erhalten werden, die anhand des Modells klassifiziert werden können, wurde

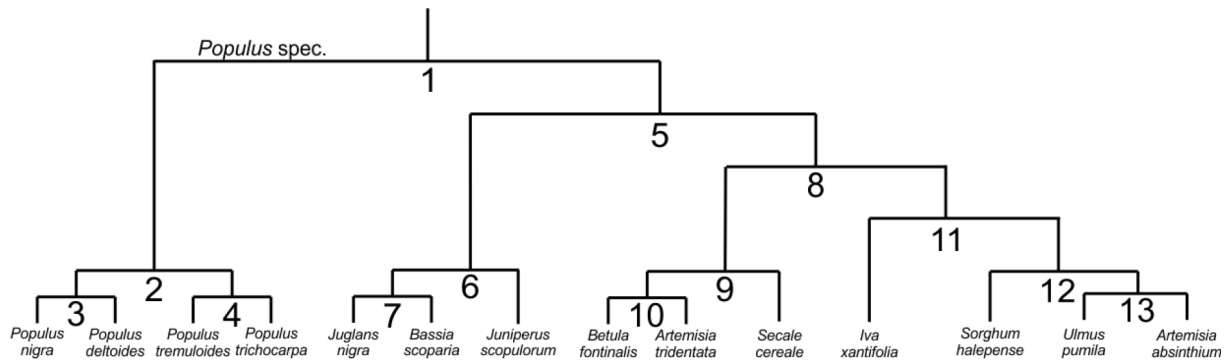
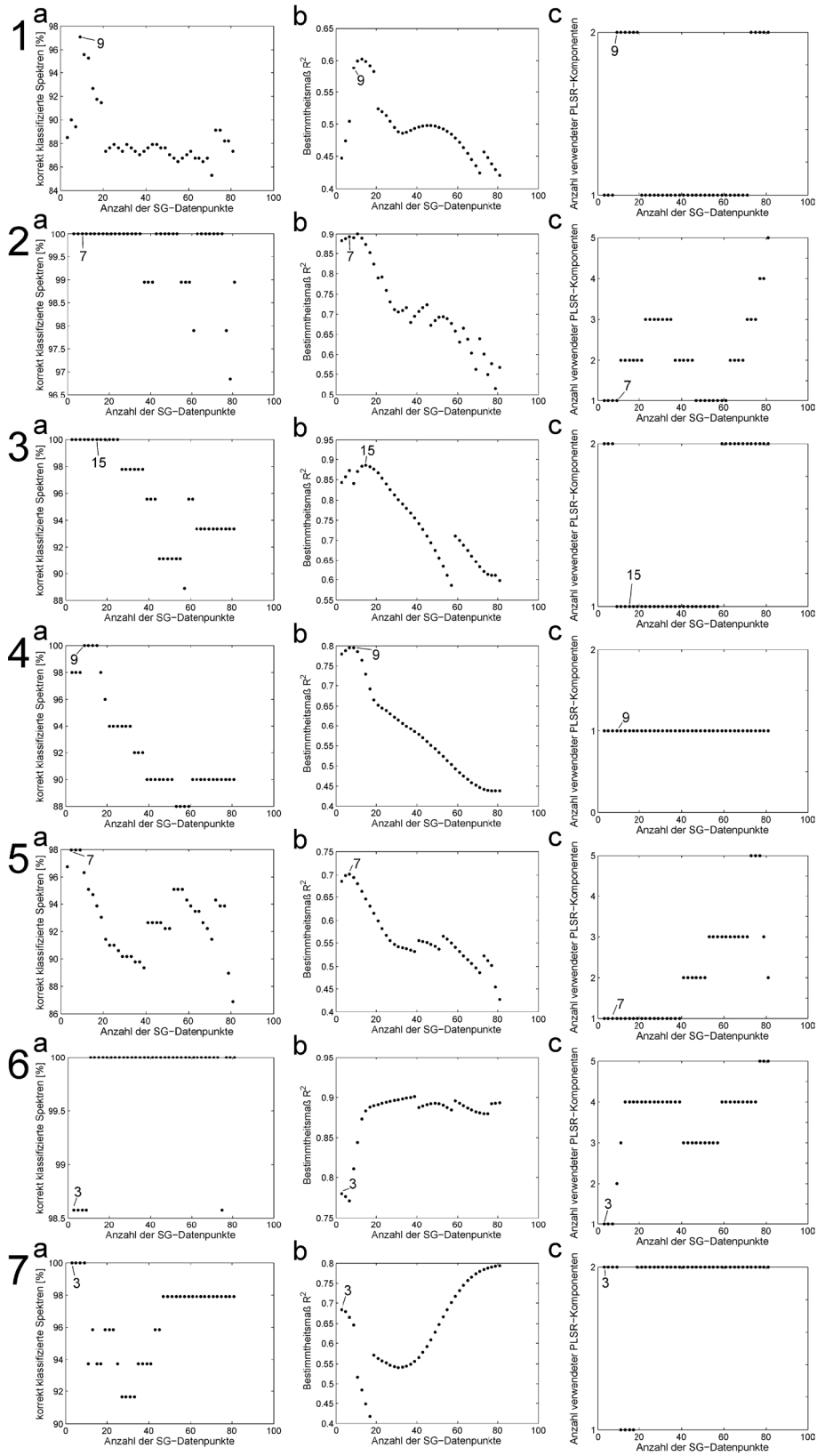


Abbildung 7.2: Varianzbasiertes, hierarchisches Klassifizierungsmodell für die Klassifizierung von FTIR-Spektren einzelner Pollenkörner von kommerziell erhältlichen Pollenproben. Die Optimierung der Anzahl an SG-Datenpunkten für die nummerierten Knotenpunkte kann der folgenden Abbildung 7.3 entnommen werden.

das Klassifizierungsmodell nicht an taxonomischen Verwandtschaftsverhältnissen, sondern an der jeweiligen Varianz der Daten an den Knotenpunkten orientiert. Um den Einfluss reiner Streuartefakte und den von Spektralbereichen mit schlechtem Signal-zu-Rausch-Verhältnis^[165] zu verringern, wurden im Klassifizierungsmodell nur die Daten im Spektralbereich von 1200 bis 2000 cm^{-1} verwendet. Abbildung 7.2 zeigt das varianzbasierte Klassifizierungsmodell, welches als Grundlage für die Optimierung diente. Es wurde an jedem Knotenpunkt eine Optimierung der Anzahl an SG-Datenpunkten bei der Bildung der zweiten Ableitung kombiniert mit einer EMSC-Korrektur der Spektren^[163] durchgeführt. Dazu wurde jeweils für jede verwendete Anzahl an SG-Datenpunkten ein PLSR-Modell zur Unterscheidung der jeweiligen Klassen aufgebaut. Für die Bestimmung der optimalen Anzahl an Datenpunkten waren drei Parameter dieses Modells relevant:

- (i) Zum einen wurde auf Basis des PLSR-Modells eine Zuordnung der Spektren zu den jeweiligen Klassen durchgeführt und somit der Anteil an *korrekt klassifizierten Spektren* berechnet.
- (ii) Zum anderen wurde bei dieser Zuordnung jeweils das *Bestimmtheitsmaß*, das sich aus dem Klassifizierungsfehler ergab, ermittelt.
- (iii) Des Weiteren wurde beim Aufbau des PLSR-Modells jeweils anhand eines χ^2 -Testes eine *optimale Anzahl an PLSR-Komponenten* für das jeweilige Modell bestimmt.

Die optimale Anzahl an SG-Datenpunkten wurde nun anschließend durch denjenigen Wert festgelegt, bei dem eine minimale Anzahl an PLSR-Komponenten für eine möglichst hohe Anzahl an korrekt klassifizierten Spektren bei möglichst hohem Bestimmtheitsmaß benötigt wird. Die Abbildung 7.3 zeigt für jeden Knotenpunkt im Klassifizierungsmodell (siehe Abbildung 7.2) die Auftragungen der drei relevanten Parameter (a bis c) in Abhängigkeit der



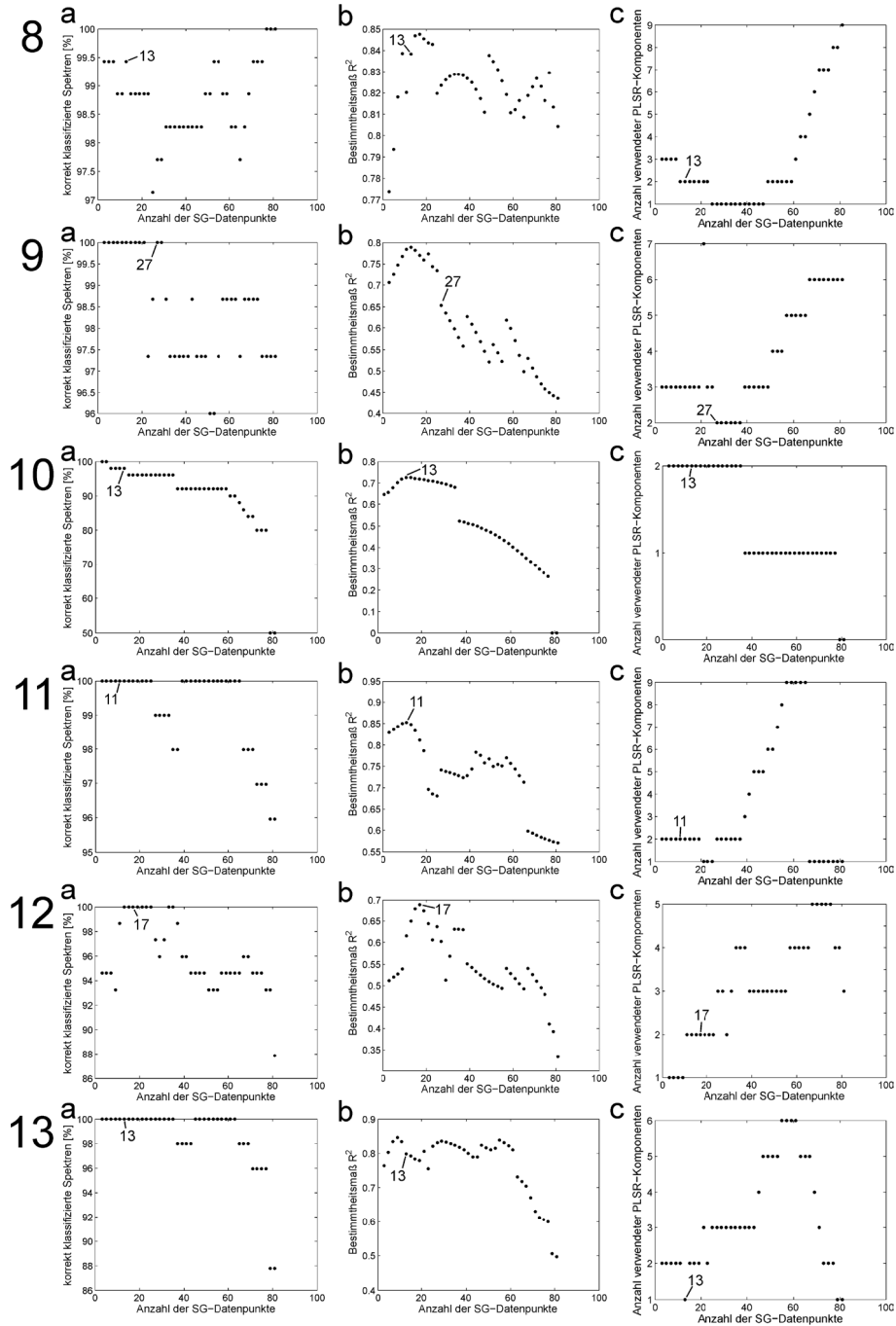


Abbildung 7.3: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte 1 bis 13 des hierarchischen Klassifizierungsmodells in Abbildung 7.2. Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt.

Anzahl der verwendeten SG-Datenpunkte und die jeweils als optimal identifizierte Anzahl wurde in den Graphen gekennzeichnet. Diese ist für alle Knotenpunkte relativ niedrig, was darauf hindeutet, dass für die Klassifizierung vor allem relativ schmale Absorptionsbanden wichtig sind. Um die Möglichkeit der hierarchischen Klassifizierung der Spektren zu überprüfen, wurden anschließend PLSR-Modelle von zweifach abgeleiteten und EMSC-korrigierten Spektren trainiert. Bei der zweiten Ableitung wurde jeweils an jedem Datenpunkt die optimierte Anzahl an SG-Datenpunkten verwendet. Die Regressionskoeffizienten des hierarchischen Modells sind in Abbildung 7.4 gezeigt. Der Bereich zwischen 1800 und 2000 cm^{-1} beeinflusst diese relativ wenig. Da in diesem Spektralbereich keine Absorptionsbanden auftreten, kann daraus geschlossen werden, dass der Einfluss des Streubeitrages auf die Modelle und damit auf eine erfolgreiche Klassifizierung relativ gering ist. Durch das Verwenden einer optimalen Anzahl an SG-Datenpunkten bei der zweiten Ableitung können also für die Klassifizierung relevante Absorptionsbeiträge der FTIR-Spektren stärker

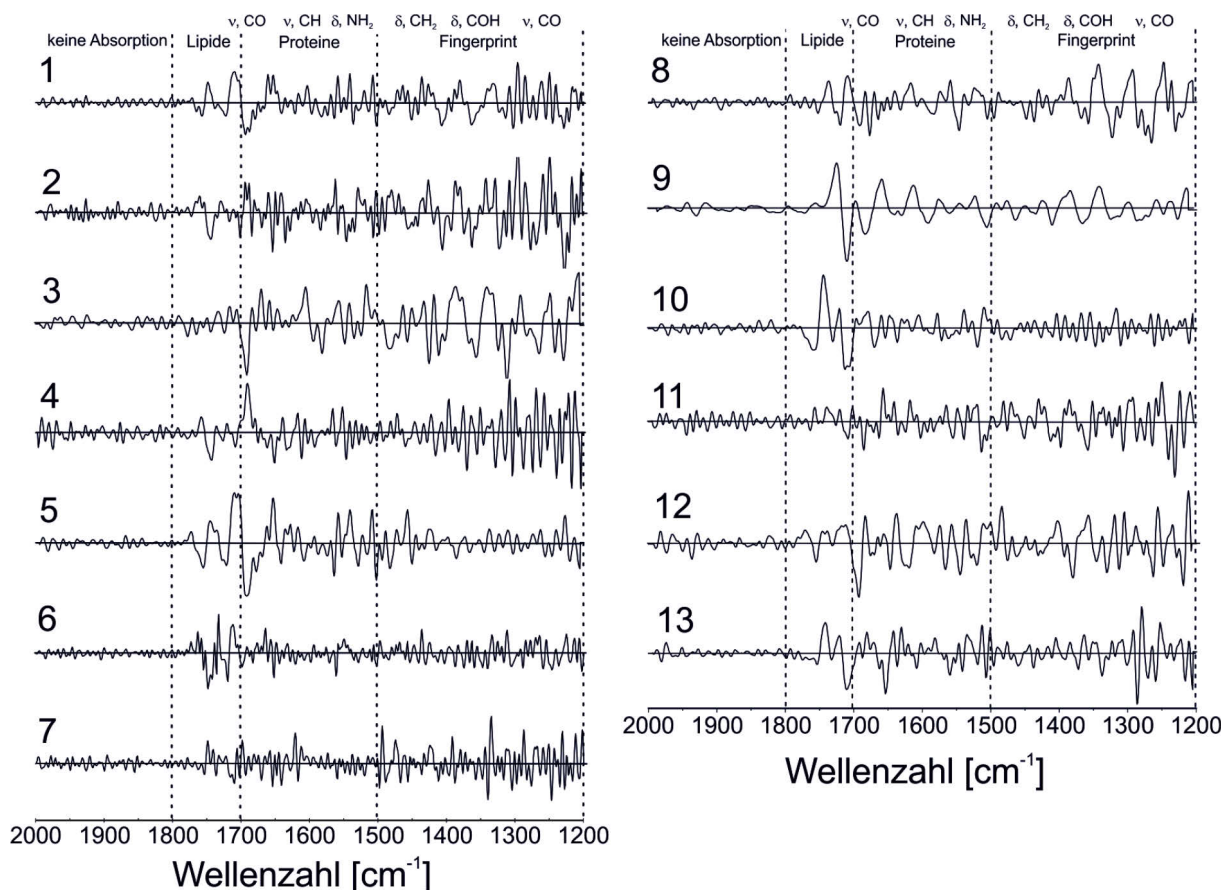


Abbildung 7.4: Regressionskoeffizienten der Knotenpunkte 1 bis 13 des hierarchischen Klassifizierungsmodells (siehe Abbildung 7.2), welches durch Training mit 140 Pollenspektren erhalten wurde.

gewichtet und der Einfluss von oszillierenden Untergrundbanden, die auf Streubeiträge zurückgeführt werden können, verringert werden.

Betrachtet man die einzelnen Regressionskoeffizienten, ist erkennbar, dass die Unterscheidungen an den Knotenpunkten auf Varianzen verschiedener spektraler Bereiche basieren. Die Klassifizierung am Knotenpunkt 5 (siehe Abbildung 7.2) ist dabei primär durch die Carbonylstreckschwingungen von Lipiden beeinflusst (Abbildung 7.4, fünfte Spur links). Dies spricht dafür, dass die Pollenkörner der Arten *Juglans nigra*, *Bassia scoparia* und *Juniperus scopulorum* Ähnlichkeiten in Lipiden aufweisen, anhand derer sie von anderen Pollenkörnern unterschieden werden können. Ein ähnlich großer Einfluss dieses Spektralbereichs ist an den Knotenpunkten 9 und 10 (siehe Abbildung 7.2) detektierbar (Abbildung 7.4, zweite und dritte Spur rechts). Auch die Pollenkörner von *Secale cereale*, *Artemisia tridentata* und *Betula fontinalis* unterscheiden sich demnach untereinander vor allem in ihren Lipidbestandteilen. An anderen Knotenpunkten zeigen die Regressionskoeffizienten einen geringeren Einfluss dieses Spektralbereichs und die Differenzierung erfolgt v.a. anhand von Banden im Fingerprintbereich zwischen 1200 und 1500 cm^{-1} . Dies gilt beispielsweise für die Knotenpunkte 4 (Abbildung 7.4, vierte Spur links) und 8 (Abbildung 7.4, erste Spur rechts). Da in diesem Bereich die Beiträge verschiedener Schwingungen und Moleküle überlagern, ist eine genaue Identifizierung der Unterschiede der Pollenkörner, die anhand dieser Knotenpunkte differenziert werden, nicht möglich. Die breiten Amidbanden im Spektralbereich zwischen 1500 und 1700 cm^{-1} beeinflussen die meisten Regressionskoeffizienten (Abbildung 7.4) vergleichsweise wenig, obwohl sie in den Spektren hohe Intensitäten aufweisen (Abbildung 7.1). Dies wird

identifiziert als														
Art der getesteten Spektren	<i>Bassia scoparia</i>	<i>Artemisia absinthium</i>	<i>Artemisia tridentata</i>	<i>Iva xanthifolia</i>	<i>Betula fontinalis</i>	<i>Juniperus scopulorum</i>	<i>Juglans nigra</i>	<i>Secale cereale</i>	<i>Sorghum halepense</i>	<i>Populus deltoides</i>	<i>Populus nigra</i>	<i>Populus tremuloides</i>	<i>Populus trichocarpa</i>	<i>Ulmus pumila</i>
<i>Bassia scoparia</i>	9(60%)	2(13%)						2(13%)	1(7%)					1(7%)
<i>Artemisia absinthium</i>	15(100%)													
<i>Artemisia tridentata</i>		13(87%)		1(7%)			1(7%)							
<i>Iva xanthifolia</i>	2(13%)	12(80%)											1(7%)	
<i>Betula fontinalis</i>			14(93%)				1(7%)							
<i>Juniperus scopulorum</i>				12(100%)										
<i>Juglans nigra</i>	3(17%)					10(83%)								
<i>Secale cereale</i>						15(100%)								
<i>Sorghum halepense</i>								15(100%)						
<i>Populus deltoides</i>								1(10%)	9(90%)					
<i>Populus nigra</i>		6(40%)					4(27%)			5(33%)				
<i>Populus tremuloides</i>	1(7%)	2(13%)	2(13%)		2(13%)						8(53%)			
<i>Populus trichocarpa</i>									1(7%)		2(14%)	12(80%)		
<i>Ulmus pumila</i>		2(14%)						3(21%)						9(64%)

Abbildung 7.5: Ergebnisse der PLSR-Klassifizierung von 199 Spektren kommerziell erhältlicher Pollenproben nach hierarchischem Klassifizierungsmodell (siehe Abbildung 7.2), welches durch Training mit 140 Pollenspektren erhalten wurde.

durch die relativ geringe Anzahl an SG-Datenpunkten bei der zweiten Ableitung der Spektren erreicht und zeigt, dass die Proteine in den Pollenkörnern bei der Klassifizierung keinen erheblichen Faktor darstellen. Das hierarchische Modell (siehe Abbildung 7.2) und die Regressionskoeffizienten (siehe Abbildung 7.4) wurden anschließend dazu verwendet, Spektren der Proben zu klassifizieren, die nicht für das Trainieren des PLSR-Modells verwendet wurden. Die Ergebnisse der Klassifizierung sind in Abbildung 7.5 gezeigt. Insgesamt werden ca. 80% der Spektren der jeweils korrekten Art zugeordnet. Damit sind in vielen Fällen tatsächlich reproduzierbare und spezifische Pollendaten in den Spektren vorhanden und für die Klassifizierung nutzbar. Allerdings werden Spektren einzelner Arten, beispielsweise die von *Populus nigra*, *Populus tremuloides* und *Bassia scoparia*, im Klassifizierungsmodell vermehrt anderen Arten zugeordnet. Dies ist auf eine falsche Zuordnung der Spektren an einzelnen Knotenpunkten des Modells (siehe Abbildung 7.2) zurückzuführen. Der Grund für diese Fehlklassifizierungen liegt vermutlich in der Überlagerung der Absorptionsbanden durch Streuartefakte. Die Fehlklassifizierung ist jedoch in vielen Fällen systematisch und nicht willkürlich, was daran ersichtlich ist, dass mehrere Spektren einer Art jeweils derselben falschen Klasse im hierarchischen Modell zugeordnet werden (Abbildung 7.5). Dies deutet darauf hin, dass trotz der fehlerhaften Zuordnung jeweils ähnliche Spektralbereiche der fehlklassifizierten Spektren vorliegen, die jedoch an einzelnen Knotenpunkte in den PLSR-Modellen nicht für eine korrekte Klassifizierung ausreichend gewichtet werden. Um dies auszugleichen, wurden zusätzliche Knotenpunkte im Modell erstellt, die zur Kontrolle häufiger Fehleinschätzungen dienen sollen. Abbildung 7.6 zeigt ein hierarchisches Modell, bei dem sieben zusätzliche Knotenpunkte (in römischen Zahlen gekennzeichnet) eingefügt wurden.

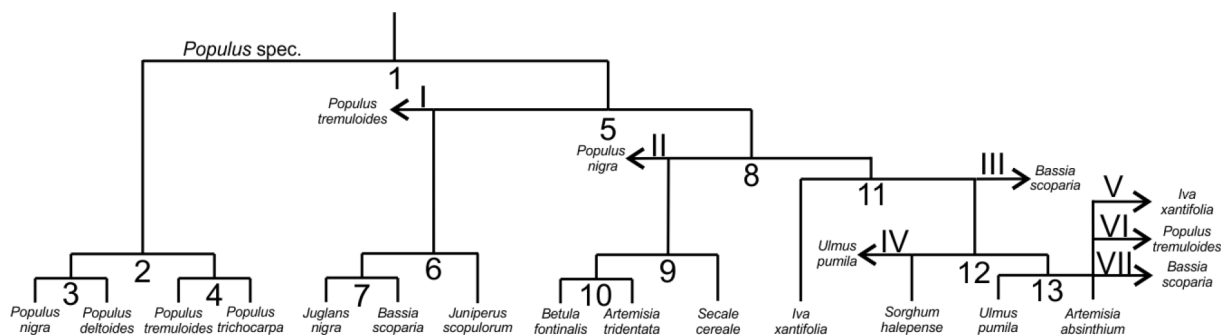


Abbildung 7.6: Optimiertes, hierarchisches Klassifizierungsmodell für die Klassifizierung von FTIR-Spektren einzelner Pollenkörner von kommerziell erhältlichen Pollenproben. Es wurden zusätzlich zum hierarchischen Modell in Abbildung 7.2 sieben Knotenpunkte eingefügt (I-VII), die der Kontrolle systematischer Fehlklassifizierung dienen. Die Optimierung der Anzahl an SG-Datenpunkten für die zusätzlichen Knotenpunkte können der folgenden Abbildung 7.7 entnommen werden.

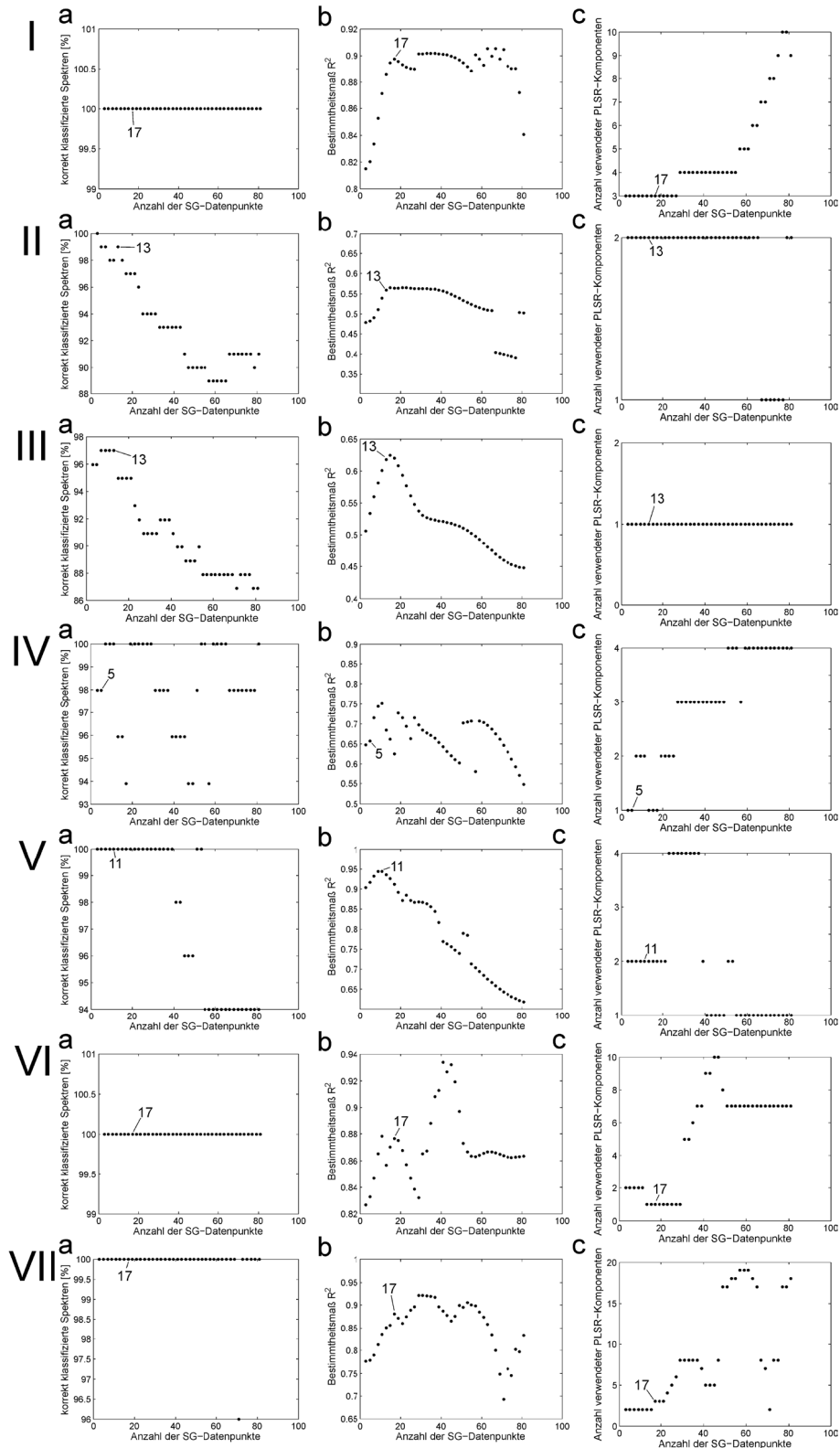


Abbildung 7.7: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte I bis VII des hierarchischen Klassifizierungsmodells in Abbildung 7.6. Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt.

Auch für diese wurde eine Optimierung der Anzahl an SG-Datenpunkten durchgeführt, (siehe Abbildung 7.7) welche erneut relativ niedrige Werte hervorbringt. Dies deutet darauf hin, dass auch diese Unterscheidungen anhand von relativ schmalen Absorptionsbanden erfolgen. Unter Verwendung der optimierten Anzahl an SG-Datenpunkten bei der zweiten Ableitung wurden anschließend PLSR-Modelle der zusätzlichen Knotenpunkte trainiert. Auch die Regressionskoeffizienten dieser Modelle (Abbildung 7.8) sind primär durch Absorptionsbanden der FTIR-Spektren und nicht durch den Spektralbereich zwischen 1800 und 2000 cm^{-1} beeinflusst. Die Regressionskoeffizienten der Knotenpunkte III und VII (Abbildung 7.8, dritte Spur links und dritte Spur rechts), die für eine Identifizierung von Pollenspektren der Art *Bassia scoparia* eingefügt wurden, zeigen dabei einen intensiven Einfluss der Carbonylstreckschwingung von Lipiden. Dies bestätigt, was aus dem Regressionskoeffizienten des Knotenpunktes 5 (siehe Abbildung 7.4, fünfte Spur links) geschlossen wurde: Pollenkörner der Art *Bassia scoparia* unterscheiden sich von den meisten anderen Pollenkörnern vor allem durch unterschiedliche Lipide. Die anderen Regressionskoeffizienten (Abbildung 7.8) zeigen jeweils, dass sie durch

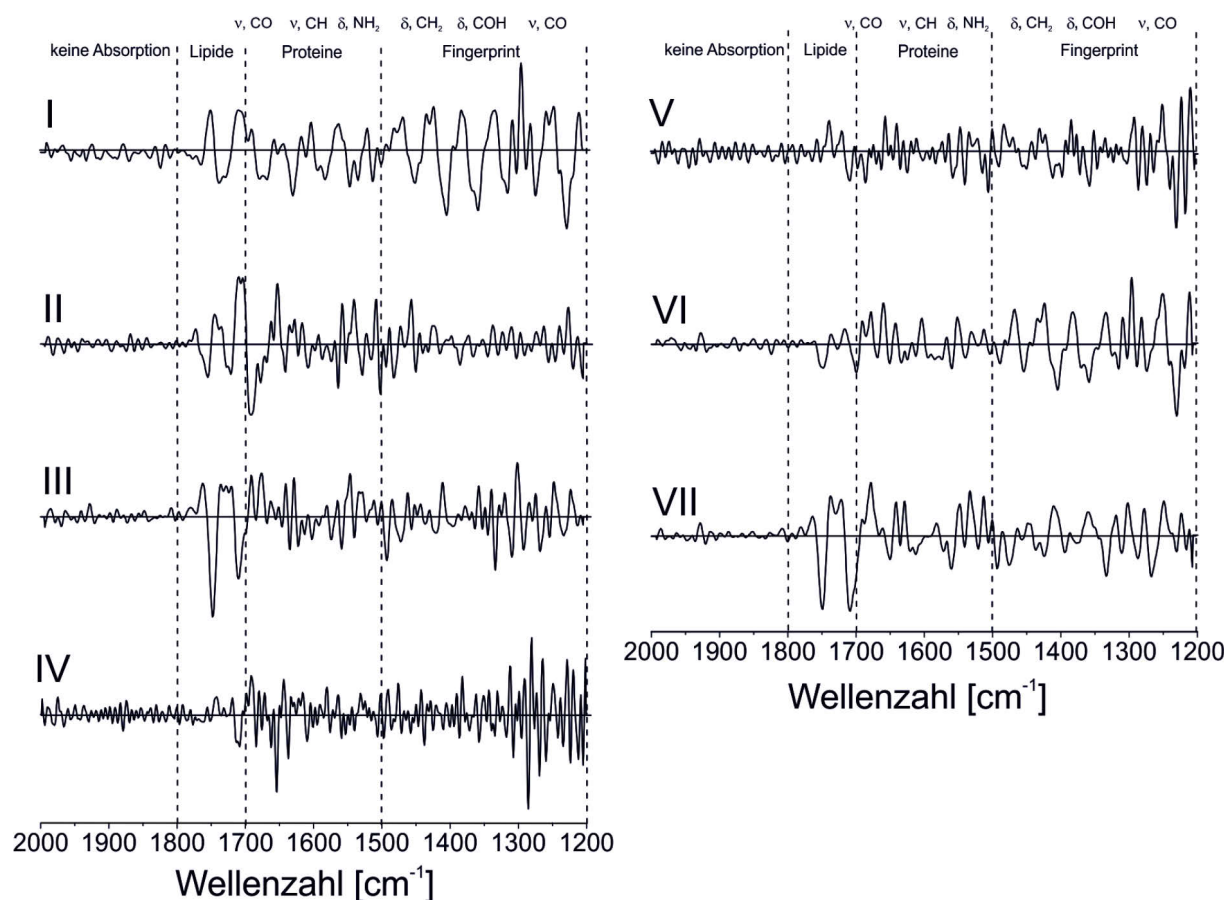


Abbildung 7.8: Regressionskoeffizienten der Knotenpunkte I bis VII des optimierten, hierarchischen Klassifizierungsmodells (siehe Abbildung 7.6), welches durch Training mit 140 Pollenspektren erhalten wurde.

verschiedene Spektralbereiche, wie z.B. durch den Fingerprintbereich zwischen 1200 und 1500 cm^{-1} beeinflusst werden. Die Klassifizierung an diesen Knotenpunkten erfolgt demnach anhand mehrerer Absorptionsbanden, für die eine Zuordnung zu spezifischen Molekülklassen schwieriger ist.

Das optimierte, hierarchische Modell wurde anschließend anhand der Spektren, die nicht für das Training der PLSR-Modelle verwendet wurden, getestet. Abbildung 7.9 zeigt die Ergebnisse der Klassifizierung. Insgesamt werden über 90% der jeweils korrekten Klasse zugeordnet und für jede Art können jeweils über 80% der Spektren korrekt klassifiziert werden. Durch das Hinzufügen zusätzlicher Knotenpunkte kann somit die Klassifizierung signifikant verbessert werden, was durch die zusätzliche Berücksichtigung von Varianzen anderer Absorptionsbanden erreicht wird. Die Ergebnisse zeigen, dass FTIR-Daten einzelner Pollenkörner, trotz der Überlagerung von Absorptions- und Streubeiträgen in den Spektren, für die taxonomische Klassifizierung verwendet werden können. Wird ein hierarchisches Modell für die Klassifizierung verwendet, an dessen Knotenpunkten jeweils eine Optimierung der Anzahl an SG-Datenpunkten beim Bilden der zweiten Ableitung durchgeführt wird, können die Absorptionsbanden der für die Klassifizierung relevanten Spektralbereiche hervorgehoben werden. Die optimierte Anzahl an SG-Datenpunkten ist dabei relativ gering, was dazu führt, dass sowohl der Einfluss von oszillierenden Untergrundbanden, als auch der von Amidbanden auf die Klassifizierung minimiert wird. Die Fehlklassifizierung einzelner Spektren durch den Einfluss von Streubeiträgen kann durch zusätzliche Knotenpunkte im Modell, die auf die Differenzierung spezifischer Klassenunterschiede ausgerichtet sind, ausgeglichen werden.

identifiziert als														
Art der getesteten Spektren	<i>Bassia scoparia</i>	<i>Artemisia absinthium</i>	<i>Artemisia tridentata</i>	<i>Iva xanthifolia</i>	<i>Betula fontinalis</i>	<i>Juniperus scopulorum</i>	<i>Juglans nigra</i>	<i>Secale cereale</i>	<i>Sorghum halepense</i>	<i>Populus deltoides</i>	<i>Populus nigra</i>	<i>Populus tremuloides</i>	<i>Populus trichocarpa</i>	<i>Ulmus pumila</i>
<i>Bassia scoparia</i>	12(80%)	1(7%)							1(7%)					1(7%)
<i>Artemisia absinthium</i>		15(100%)												
<i>Artemisia tridentata</i>			13(87%)		1(7%)		1(7%)							
<i>Iva xanthifolia</i>				14(93%)									1(7%)	
<i>Betula fontinalis</i>					14(93%)		1(7%)							
<i>Juniperus scopulorum</i>						12(100%)								
<i>Juglans nigra</i>	3(17%)					10(83%)								
<i>Secale cereale</i>							15(100%)							
<i>Sorghum halepense</i>								15(100%)						
<i>Populus deltoides</i>								1(10%)	9(90%)					
<i>Populus nigra</i>										15(100%)				
<i>Populus tremuloides</i>				2(13%)								13(87%)		
<i>Populus trichocarpa</i>									1(7%)		2(13%)	12(80%)		
<i>Ulmus pumila</i>		2(13%)						1(7%)						12(80%)

Abbildung 7.9: Ergebnisse der PLSR-Klassifizierung von 199 Spektren kommerziell erhältlicher Pollenproben nach optimiertem hierarchischem Klassifizierungsmodell (siehe Abbildung 7.6), welches durch Training mit 140 Pollenspektren erhalten wurde.

7.2 Wiedergabe taxonomischer Verwandtschaftsverhältnisse aus den FTIR-Daten einzelner Pollenkörner

Im vorigen Abschnitt wurde gezeigt, dass es möglich ist, durch FTIR-Spektroskopie im Transmissionsmodus spezifische, reproduzierbare Fingerabdrücke einzelner Pollenkörner zu erhalten, um diese für die Klassifizierung von Pollenproben zu verwenden. Als nächster Schritt soll nun untersucht werden, ob anhand charakteristischer Fingerabdrücke auch taxonomische Verwandtschaftsverhältnisse reproduziert werden können. Es wurden deshalb jeweils ca. 20 Spektren von 77 Pollenproben 33 verschiedener Arten von neun Gattungen der Ordnungen *Fagales* und *Coniferales* (siehe Tabelle 3.2) aufgenommen. Aufgrund der hohen Anzahl analysierter Proben können bei dieser Untersuchung jeweils Spektren von Pollenkörnern unterschiedlicher Individuen für das Trainieren und das Testen des Klassifizierungsmodells verwendet werden. Auf diese Weise ist es möglich zu überprüfen, ob die spezifischen

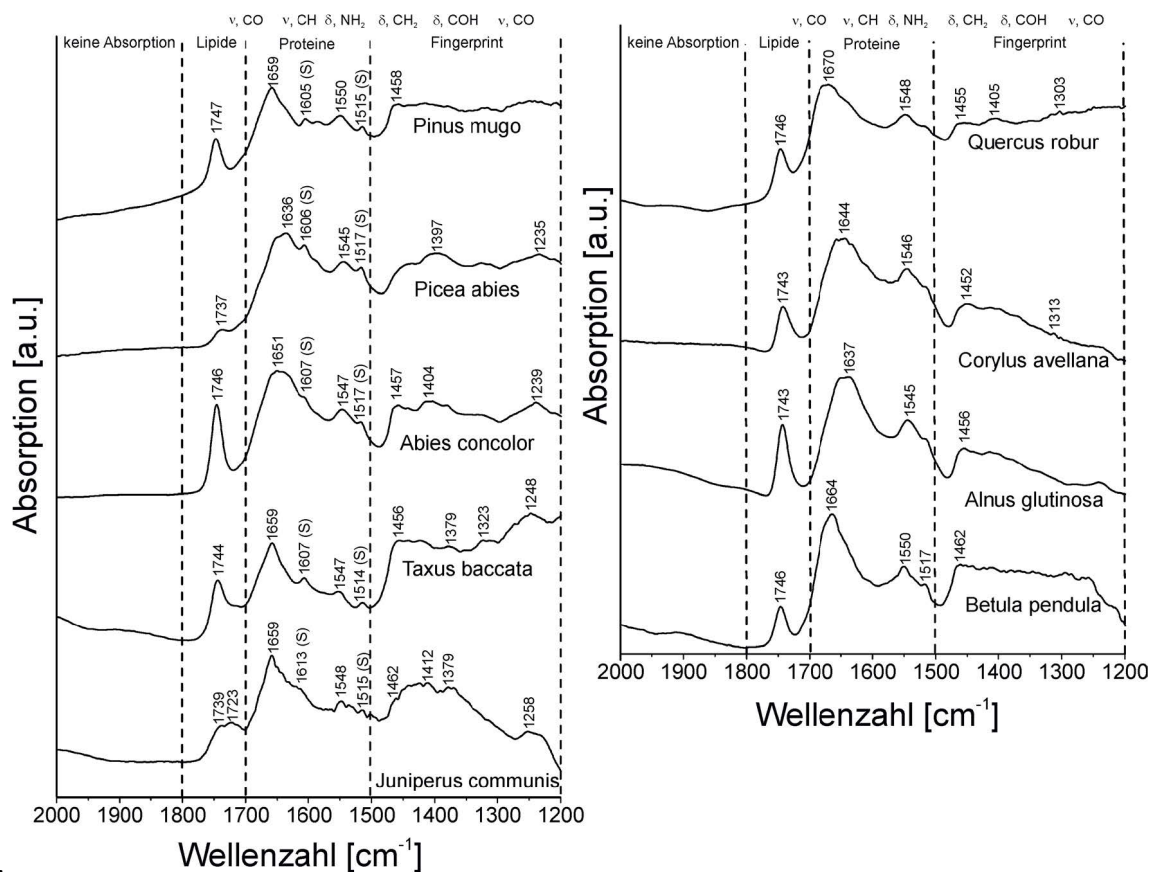


Abbildung 7.10: Ausgewählte FTIR-Spektren einzelner Pollenkörner der Ordnungen *Coniferales* (links) und *Fagales* (rechts). Es sind Spektralbereiche gekennzeichnet, in denen charakteristische Banden von Lipiden und Proteinen detektiert werden können. Sporopollenin-spezifische Banden sind zusätzlich mit einem S gekennzeichnet.

Fingerabdrücke der Pollenproben charakteristisch für die jeweiligen Proben und auch für die taxonomischen Gruppen sind. Abbildung 7.10 zeigt Spektren, die aufgrund relativ weniger, offensichtlicher Streuartefakte ausgewählt wurden, um die Absorptionsbanden der FTIR-Daten zu vergleichen. Es fällt zunächst auf, dass große Variationen in der Intensität der Carbonylstreckschwingung von Lipiden bei ca. 1745 cm^{-1} beobachtet werden können. Da die Lipidzusammensetzung allerdings auch bei Pollenkörnern einer Art, bzw. einer Gattung stark variieren kann und z.B. auch durch Umwelteinflüsse^[17] bestimmt wird, ist diese Bande nicht zwangsläufig für eine Unterscheidung verschiedener taxonomischer Gruppen geeignet. Charakteristisch für Pollenkörner der Familie *Pinaceae* ist, dass diese aus einem Pollenkörper, der *Corpus* genannt wird, und zumeist zwei Luftsäcken, den sogenannten *Sacci*, bestehen (siehe Abbildung 7.11, oben). Da es sich bei den *Sacci* um Hohlkörper handelt, die primär aus der Außenhülle von Pollenkörnern gebildet werden, ist bei ihnen der relative Sporopollenin-gehalt bedeutend größer als bei den Bestandteilen des *Corpus*. Deshalb treten in den FTIR-Spektren dieser Pollenkörner neben den Amidbanden des *Corpus* bei ca. 1655 cm^{-1} und 1547 cm^{-1} auch verstärkt solche Banden z.B. bei ca. 1606 cm^{-1} und 1516 cm^{-1} auf, die Sporopolleninbestandteilen zugeordnet werden können (siehe Abbildung 7.10, erste bis dritte Spur links).^[165] Dass diese spektralen Merkmale in ähnlicher Intensität ebenfalls in den FTIR-Spektren von *Taxus*- und *Juniperus*-Pollenkörnern auftreten (siehe Abbildung 7.10, vierte und fünfte Spur links), ist zunächst überraschend, da diese keine *Sacci* besitzen. Die Pollenkörner dieser Gattungen sind jedoch relativ klein (siehe Abbildung 7.11, unten), was dazu führt, dass

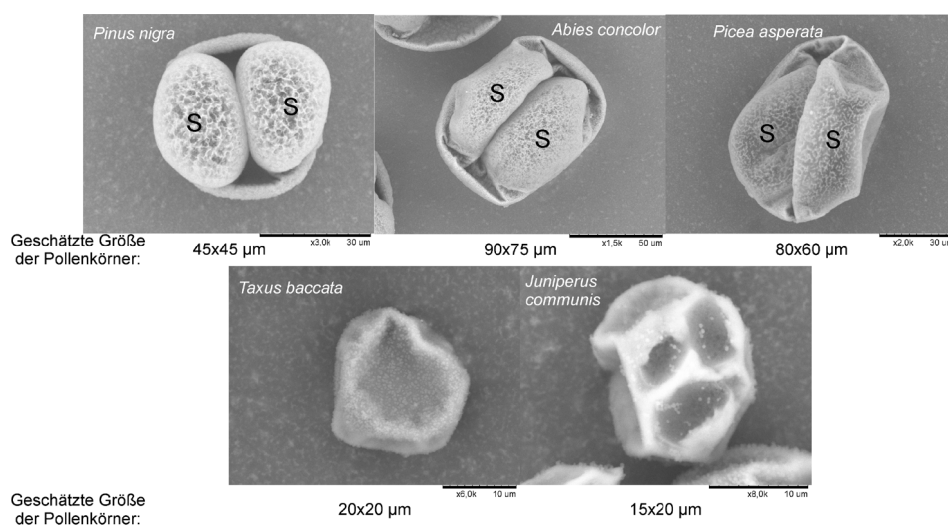


Abbildung 7.11: ESEM Aufnahmen von Pollenkörnern der Ordnung *Coniferales* und deren geschätzte Größe. Die *Sacci* der Pollenkörner der Familie *Pinaceae* sind mit einem S gekennzeichnet.

ein relativ großer Anteil des untersuchten Pollenkorns aus Außenhüllenkomponenten besteht.^[16] Relativ intensive Sporopolleninbanden in den FTIR-Spektren sind somit kennzeichnend für Spektren der Ordnung *Coniferales* und könnten für die Klassifizierung der taxonomischen Ebene der Ordnung relevant sein. Die relativ breiten Amid I- und Amid II-Banden des *Corpus* von *Pinaceae*-Pollenkörnern sind spezifische Merkmale der Spektren, um diese von den vergleichsweise schmalen Banden der Gattungen *Taxus* und *Juniperus* zu unterscheiden. Die FTIR-Spektren von *Juniperus*-Pollenkörnern werden bekanntermaßen besonders durch Streuartefakte beeinträchtigt.^[27] Betrachtet man die FTIR-Spektren der Ordnung *Fagales* (Abbildung 7.10, rechts), so ist erkennbar, dass sich diese relativ ähnlich sind. Trotz dieser Ähnlichkeit wurden aufgrund von FTIR-Spektren Kaliumbromidpresslingen verschiedener Gattungen unterschieden.^[17] Gerade bei solchen gleichartigen Spektren ist es interessant zu analysieren, inwiefern Streuartefakte die korrekte Klassifizierung beeinträchtigen. Vermutlich ähneln sich diese Spektren zusätzlich in ihren Streuartefakten. Dies kann daraus geschlossen werden, dass die Pollenkörner der Ordnung *Fagales* ähnliche Größen und die der Familie *Betulaceae*, also der Gattungen *Alnus*, *Corylus* und *Betula*, sogar eine gleichartige Morphologie (Abbildung 7.12) aufweisen. Abbildung 7.13 zeigt das hierarchische Klassifizierungsmodell, das der PLSR-Klassifizierung durch Optimierung der SG-Datenpunkte zugrunde liegt. Dieses orientiert sich an taxonomischen Verwandtschaftsverhältnissen der Pollenproben. Da jedoch die Pollenkörner der Familie *Pinaceae* sich sowohl durch deren Absorptionsspektren (siehe Abbildung 7.10) als auch durch Morphologie und Größe (siehe Abbildung 7.11) von denen der Gattungen *Taxus* und *Juniperus* unterscheiden, wurde im ersten Knotenpunkt des Modells von einer streng taxonomischen Klassifizierung der Ordnungen abgesehen. Es wurde an diesem Knotenpunkt eine Differenzierung der Spektren in solche der Pollenkörner der Familie *Pinaceae* und solche der anderen Familien angestrebt. Eine Unterscheidung der verschiedenen taxonomischen Ordnungen ist trotzdem anhand dieses Modells möglich, nämlich durch Kombination des ersten und vierten Knoten-

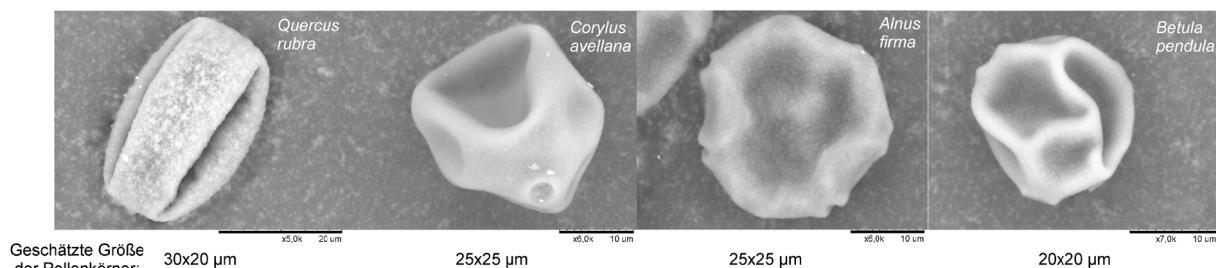


Abbildung 7.12: ESEM Aufnahmen von Pollenkörnern der Ordnung *Fagales* und deren geschätzte Größe.

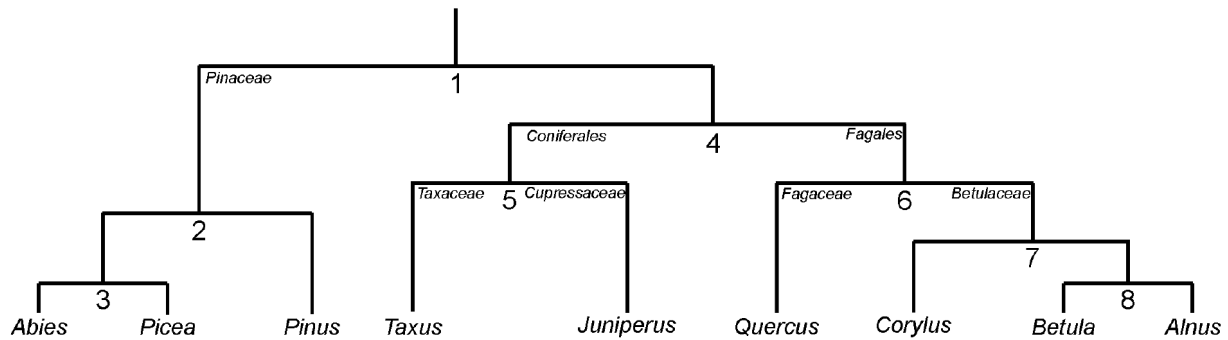
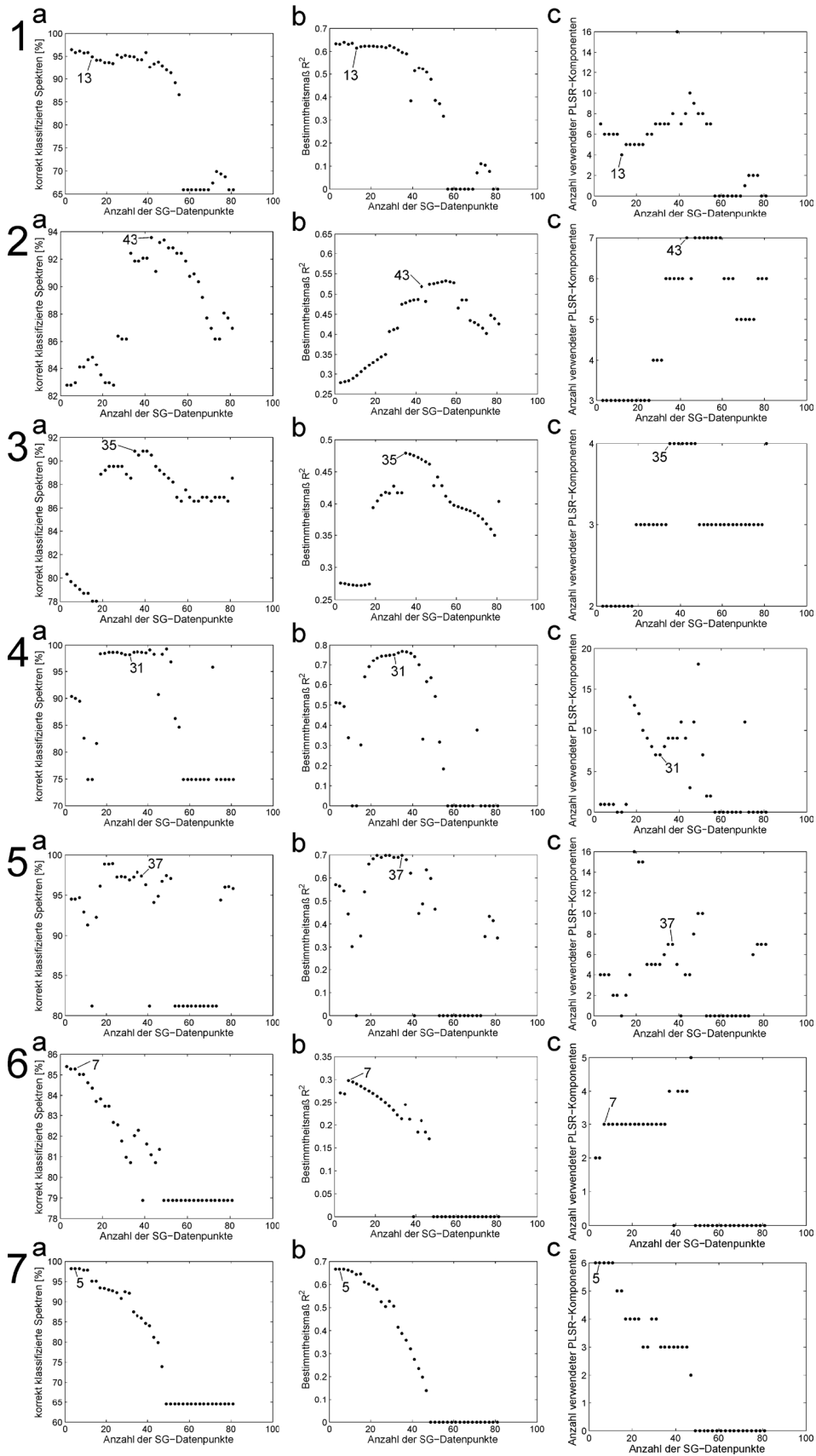


Abbildung 7.13: Hierarchisches Klassifizierungsmodell für die taxonomische Klassifizierung von FTIR-Spektren einzelner Pollenkörner von Pollenproben aus dem Feld. Die Optimierung der Anzahl an SG-Datenpunkten für die nummerierten Knotenpunkte kann der folgenden Abbildung 7.14 entnommen werden.

punktes (Abbildung 7.13). Wie bei der Untersuchung der kommerziell erhältlichen Pollenproben (siehe Abschnitt 7.1) wurde nun an jedem Knotenpunkt des Modells eine Optimierung der Anzahl an SG-Datenpunkten beim Bilden der zweiten Ableitung in Kombination mit einer EMSC-Korrektur der spektralen Fingerabdrücke im Spektralbereich zwischen 1200 und 2000 cm^{-1} durchgeführt. Abbildung 7.14 zeigt die Auftragungen der Anzahl an korrekt klassifizierten Spektren (a), des Bestimmtheitsmaßes R^2 (b) und der Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an verwendeten SG-Datenpunkten für jedes PLSR-Modell des hierarchischen Modells (Abbildung 7.13). Es wurde jeweils die Anzahl als optimal identifiziert, bei der die maximale Menge Spektren mit möglichst hohem Bestimmtheitsmaß unter Verwendung einer möglichst geringen Anzahl an PLSR-Komponenten korrekt klassifiziert wurden. Die optimierten Werte in Abbildung 7.14 zeigen, dass für die Unterscheidung von Gattungen der Ordnung *Coniferales* an den Knotenpunkten 2, 3 und 5 (siehe Abbildung 7.13) relativ hohe Werte von 30 bis 40 erhalten werden, während bei der Differenzierung von Gattungen der Ordnung *Fagales* an den Knotenpunkten 6, 7 und 8 (siehe Abbildung 7.13) relativ niedrige Werte zwischen 5 und 15 als optimal identifiziert werden. Dies spricht dafür, dass für die Differenzierung der Spektren verschiedener Gattungen der Ordnung *Coniferales* relativ breite Banden, wie z.B. Amidbanden, relevant sind, während für die der Ordnung *Fagales* schmale Banden wichtiger sind. Das Hervorheben relativ schmaler Banden zur Klassifizierung könnte auch darauf hinweisen, dass breite, oszillierende Streuartefakte in den Spektren die Klassifizierung behindern. Ein weiterer Hinweis hierauf ist, dass an den Knotenpunkten 6 und 8 (siehe Abbildung 7.14 6b und 8b) allgemein relativ niedrige Werte für das Bestimmtheitsmaß erhalten werden. Deshalb ist es möglich, dass auch bei optimierter Anzahl an SG-Datenpunkten eine Identifizierung von *Quercus*-Pollenspektren, sowie eine Differenzierung zwischen *Alnus*- und *Betula*-Pollenspektren für viele der Spektren scheitern.



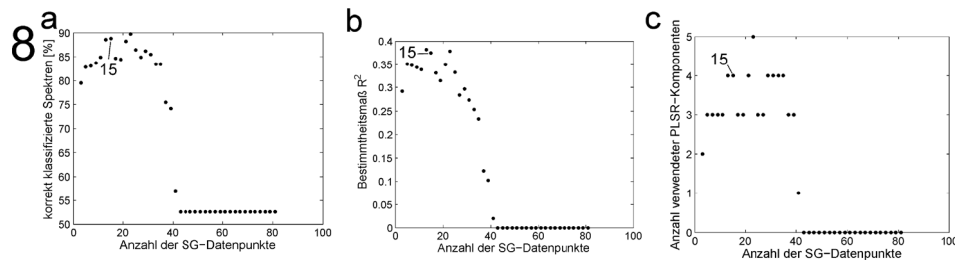


Abbildung 7.14: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte 1 bis 8 des hierarchischen Klassifizierungsmodells (siehe Abbildung 7.13). Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt.

Um den Erfolg der Klassifizierung zu testen, ist es notwendig, ein Klassifizierungsmodell mit solchen Daten zu testen, die nicht für das Erstellen des jeweiligen Modells verwendet wurden. Deshalb wurde der Datensatz geteilt und unter Verwendung einer Hälfte der Daten für jeden Knotenpunkt ein PLSR-Modell mit zweifach abgeleiteten, EMSC-korrigierten Spektren aufgebaut. Die Regressionskoeffizienten des hierarchischen Modells sind in Abbildung 7.15 gezeigt. Der Spektralbereich zwischen 1800 und 2000 cm^{-1} beeinflusst die Regressionskoeffizienten relativ wenig, woraus geschlossen werden kann, dass die Klassifizierung primär durch Absorptionsbanden bestimmt wird. Die Sporopollenin-spezifischen Banden können in den Regressionskoeffizienten der Knotenpunkte 1 und 4 identifiziert werden (siehe Abbildung 7.15, erste und vierte Spur links). Dies spricht dafür, dass diese Banden charakteristisch für Pollenkörner der Ordnung *Coniferales* sind, was bereits beim Vergleich der Spektren festgestellt wurde (siehe Abbildung 7.10). Allerdings sind für die Differenzierung an diesen Knotenpunkten auch andere Banden relevant. *Pinaceae*-Pollenkörner werden am Knotenpunkt 1 (siehe Abbildung 7.13) zusätzlich anhand von Banden im Fingerprintbereich und durch Amidbanden (Abbildung 7.15, erste Spur links) klassifiziert. Für die Unterscheidung von *Taxaceae* und *Cupressaceae*-Spektren von denen der Ordnung *Fagales* an Knotenpunkt 4 (siehe Abbildung 7.13) sind zusätzlich lipidspezifische Banden (Abbildung 7.15, vierte Spur links) entscheidend. Die Unterscheidung von *Pinus*-Pollenkörnern von denen anderer Gattungen der Familie *Pinaceae* in Knotenpunkt 2 (siehe Abbildung 7.13) erfolgt anhand verschiedener Spektralbereiche und v.a. anhand von Amid- und Sporopolleninbanden (Abbildung 7.15, zweite Spur links). Im Gegensatz dazu beeinflusst die Carbonylstreckschwingung von Lipiden die Differenzierung zwischen Spektren von *Picea* und *Abies* (Abbildung 7.15, dritte Spur) an Knotenpunkt 3 (siehe Abbildung 7.13) bzw. von *Juniperus* und *Taxus* (Abbildung 7.15, erste Spur rechts) an Knotenpunkt 5 (siehe Abbildung

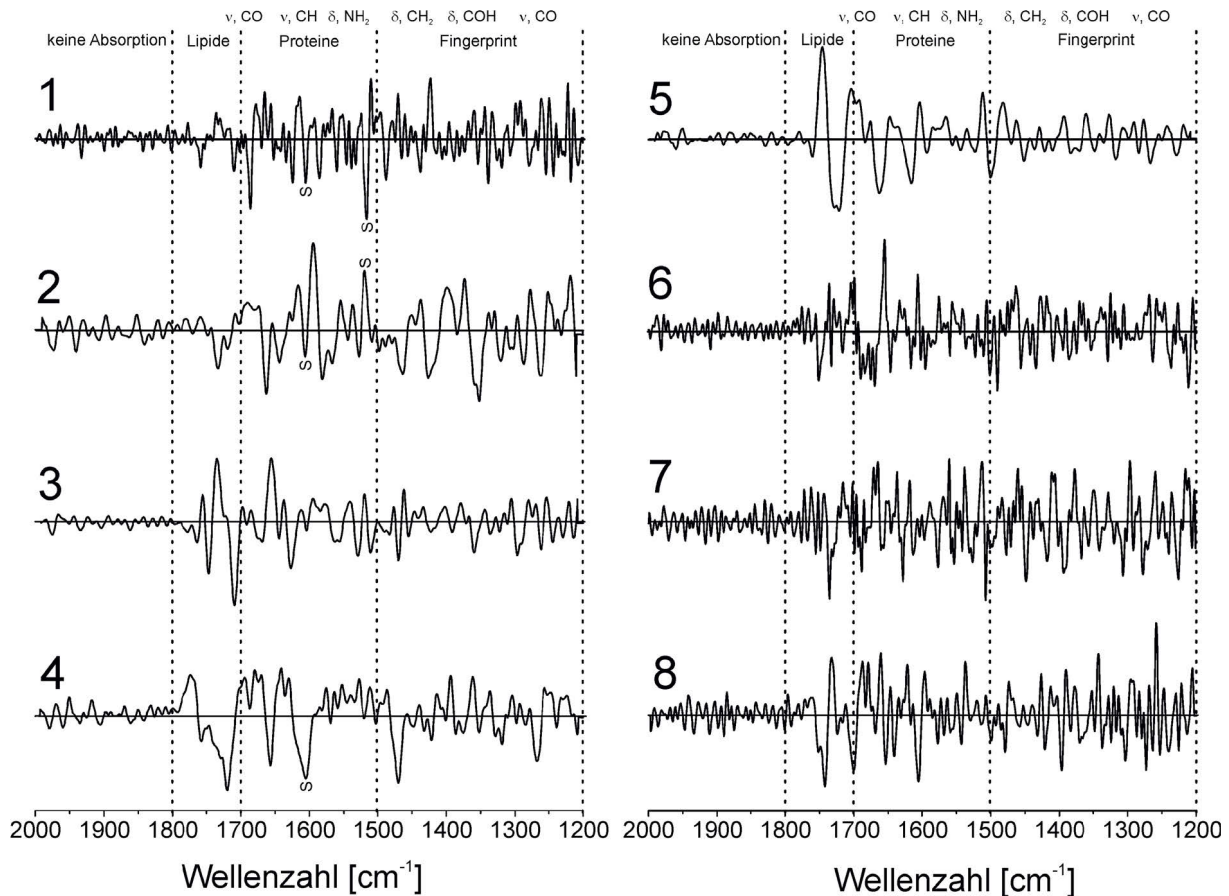


Abbildung 7.15: Regressionskoeffizienten der Knotenpunkte 1 bis 8 des hierarchischen Klassifizierungsmodells (siehe Abbildung 7.13), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde. Sporopollenin-spezifische Banden sind mit einem S gekennzeichnet.

7.13) in hohem Maße. Deutliche Unterschiede in der Intensität dieser Bande sind auch in den Spektren in Abbildung 7.10 ersichtlich. Die Pollenkörner dieser Gattungen unterscheiden sich demnach jeweils vor allem anhand von Lipiden. Für die Differenzierung verschiedener Gattungen der Ordnung *Fagales* (siehe Abbildung 7.13) sind eine Vielzahl verschiedener Banden im gesamten Spektralbereich (Abbildung 7.15, zweite bis vierte Spur rechts) relevant. Dies deutet darauf hin, dass spektrale Bestandteile verschiedener Molekülklassen zur Unterscheidung beitragen. Allerdings könnte dies auch ein Hinweis darauf sein, dass Streuartefakte die geringen spektralen Unterschiede zwischen den Gattungen überlagern und eine korrekte Klassifizierung verhindern.

Um zu überprüfen, ob eine taxonomische Klassifizierung anhand der FTIR-Daten erfolgen kann, wurde anschließend mit zweifach abgeleiteter, EMSC-korrigierten Spektren von Proben, die nicht für das Trainieren des Klassifizierungsmodells verwendet wurden, das hierarchische Modell in Abbildung 7.13 getestet. Abbildung 7.16 zeigt die Ergebnisse der Klassifizierung auf Ebene der Ordnung. Es werden 86% der Spektren der Ordnung *Coniferales*

identifiziert als		
Ordnung der getesteten Spektren	Coniferales	Fagales
Coniferales	347(86%)	56(14%)
Fagales	13(3%)	368(97%)

Abbildung 7.16: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Ordnung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach hierarchischem Klassifizierungsmodell (siehe Abbildung 7.13), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert.

und 97% der Spektren der Ordnung *Fagales* und damit insgesamt 91% aller Spektren der zutreffenden Ordnung zugewiesen. Eine inkorrekte Klassifizierung erfolgt demnach vor allem dadurch, dass spektrale Merkmale der Ordnung *Coniferales* an den Knotenpunkten 1 und 4 (siehe Abbildung 7.13) bei der Klassifizierung nicht identifiziert werden können. Dies könnte dadurch verursacht werden, dass Streuartefakte *Coniferales*-spezifische Banden, wie z.B. Sporopolleninbanden (siehe Abbildung 7.10), überlagern. Die meisten Spektren werden allerdings der korrekten Ordnung zugewiesen, weshalb geschlussfolgert werden kann, dass trotz vorhandener Streuartefakte in den FTIR-Spektren eine taxonomische Klassifizierung auf taxonomischer Ebene der Ordnung möglich ist.

Anschließend wurde anhand des hierarchischen Modells (Abbildung 7.13) eine Klassifizierung der Spektren auf taxonomischer Ebene der Gattung durchgeführt. Insgesamt werden 68% der Spektren korrekt klassifiziert (Abbildung 7.17). Die Klassifizierung ist jedoch bei den Spektren verschiedener Gattungen sehr unterschiedlich erfolgreich. Bei den Gattungen *Juniperus*, *Taxus* und *Picea* werden beispielsweise jeweils mindestens 80% der jeweils korrekten Gattung zugewiesen, während dies bei den Gattungen *Quercus* und *Betula* bei nur

identifiziert als									
Gattung der getesteten Spektren	Abies	Picea	Pinus	Juniperus	Taxus	Quercus	Corylus	Alnus	Betula
Abies	35(57%)	2 (3%)	6 (10%)			8 (13%)		10 (16%)	
Picea	7 (9%)	65 (80%)	7 (9%)					1 (1%)	1 (1%)
Pinus	6 (5%)	8 (7%)	90 (74%)			4 (3%)		12 (10%)	2 (2%)
Juniperus				54(92%)		1 (2%)			4 (7%)
Taxus			1 (1%)		66 (83%)			4 (5%)	9 (11%)
Quercus				1 (1%)	2 (2%)	31 (38%)	1 (1%)	34 (41%)	13 (16%)
Corylus						2 (2%)	98 (84%)	12 (10%)	4 (3%)
Alnus	2 (3%)		1 (1%)	1 (1%)		1 (1%)	8 (10%)	67 (84%)	
Betula	2 (2%)	1 (1%)	3 (3%)			8 (8%)	5 (5%)	53 (51%)	31 (30%)

Abbildung 7.17: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Gattung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach hierarchischem Klassifizierungsmodell (siehe Abbildung 7.13), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert.

unter 40% der Spektren der Fall ist. Da bei letzteren jeweils ein sehr großer Anteil der Spektren wie *Alnus*-Pollenspektren klassifiziert wird, lässt sich ableiten, dass Spektren der Gattungen *Alnus*, *Betula* und *Quercus* nicht unterscheidbar sind. Damit ein Spektrum als eines der Gattung *Alnus* identifiziert wird, durchläuft es das hierarchische Modell (siehe Abbildung 7.13) bis zur finalen Klassifizierung in Knotenpunkt 8. In den vorherigen Knotenpunkten werden dabei spektrale Merkmale abgefragt, die das Spektrum als das der jeweiligen anderen taxonomischen Gruppen kennzeichnet. Vermutlich werden somit Spektren anderer Gattungen, bei denen spezifische, spektrale Merkmale durch Streuartefakte gestört sind, fälschlicherweise vor allem als *Alnus*-Spektren identifiziert. Dies ist sowohl anhand der fehlerhaften Klassifizierung von Spektren der Gattungen *Betula* und *Quercus*, als auch anhand derer von Spektren der Gattungen *Corylus*, *Abies* und *Pinus* ersichtlich (Abbildung 7.17). Pollenkörner der Gattung *Corylus* können jedoch, trotz der Ähnlichkeit ihrer Absorptionsspektren (siehe Abbildung 7.10), ihrer Morphologie und ihrer Größe (siehe Abbildung 7.12) mit denen anderer Gattungen der Ordnung *Fagales*, mit hoher Wahrscheinlichkeit korrekt klassifiziert werden. Die spezifischen spektralen Merkmale dieser Gattung sind also, trotz vorhandener Streuartefakte in den Spektren, die Grundlage der Klassifizierung in Knotenpunkt 6 (siehe Abbildung 7.13).

Analog zur Klassifizierung von Spektren kommerziell erhältlicher Pollenproben (Abschnitt 7.1) wurden sechs zusätzliche Knotenpunkte im hierarchischen Modell (in römischen Zahlen in Abbildung 7.18 gekennzeichnet) ergänzt, um systematische Fehlklassifizierungen auszugleichen. Es wurden allerdings keine zusätzlichen Knotenpunkte für die Unterscheidung von Spektren der Gattungen *Alnus*, *Quercus* und *Betula* hinzugefügt, da dies nicht zu einer verbesserten Differenzierung zwischen den Spektren dieser Gattungen führte (Ergebnisse hier

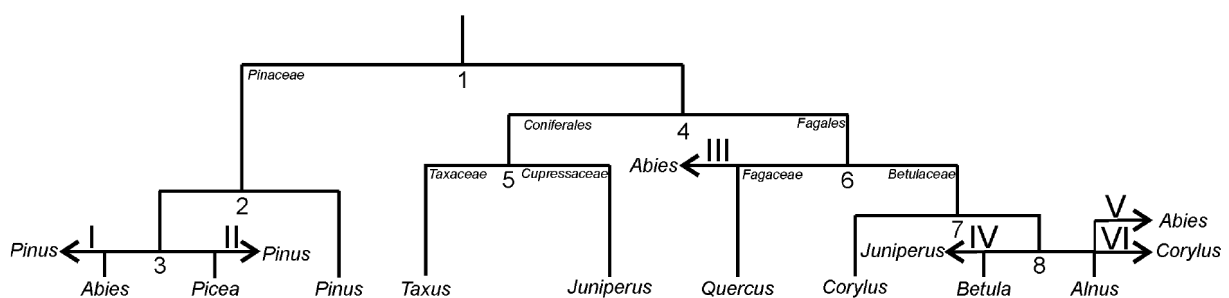


Abbildung 7.18: Optimiertes, hierarchisches Klassifizierungsmodell für die Klassifizierung von FTIR-Spektren einzelner Pollenkörner von Pollenproben aus dem Feld. Es wurden zusätzlich zum hierarchischen Modell in Abbildung 7.13 sechs Knotenpunkte eingefügt (I-VI), die der Kontrolle systematischer Fehlklassifizierung dienen. Die Optimierung der Anzahl an SG-Datenpunkten für die zusätzlichen Knotenpunkte können der folgenden Abbildung 7.19 entnommen werden.

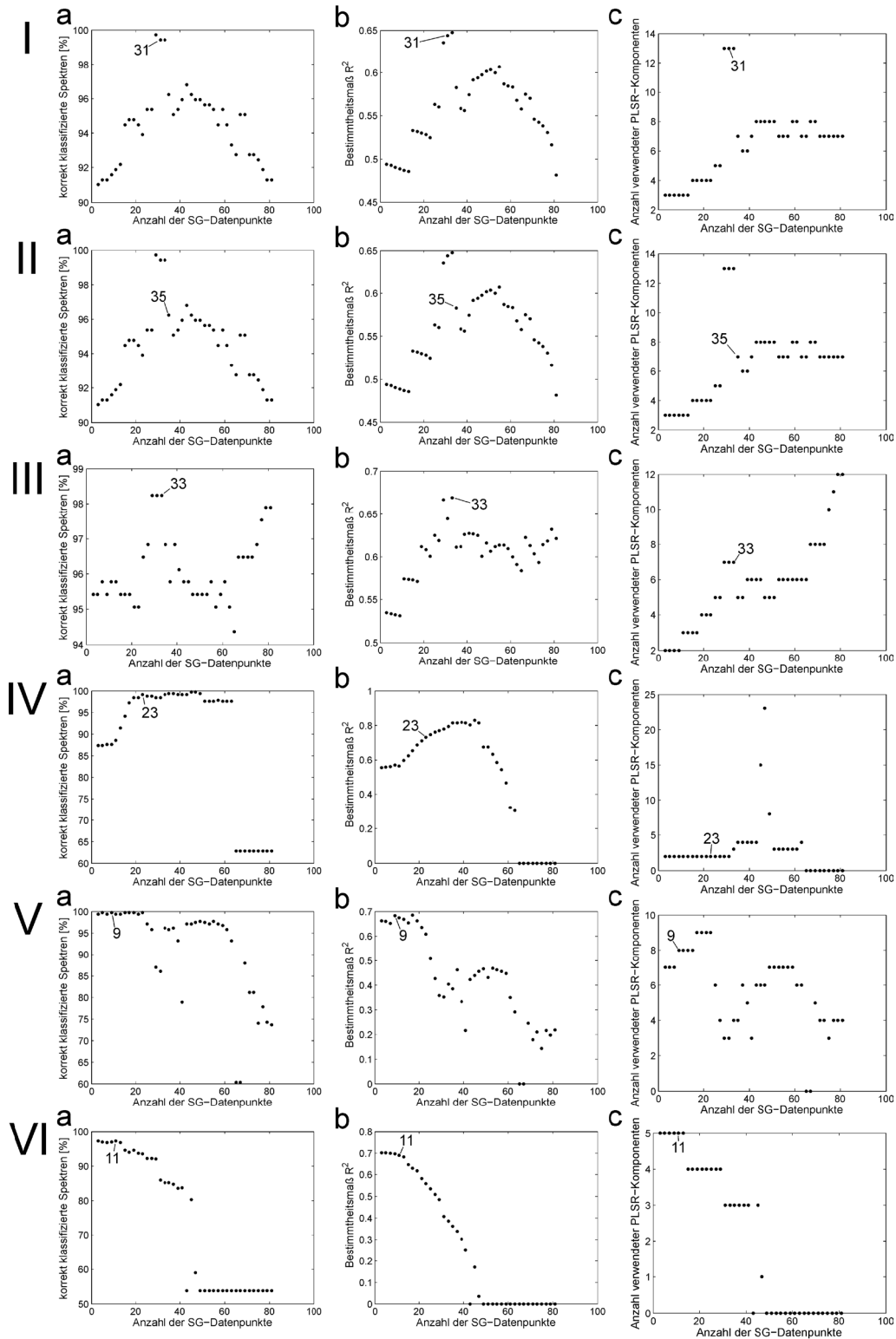


Abbildung 7.19: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte I bis VI des hierarchischen Klassifizierungsmodells in Abbildung 7.18. Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt.

nicht gezeigt). Auch für die zusätzlichen Knotenpunkte wurde eine Optimierung der Anzahl an SG-Datenpunkten beim Bilden der zweiten Ableitung in Kombination mit einer EMSC-Korrektur der Spektren durchgeführt. Die Ergebnisse dieser Optimierung sind in Abbildung 7.19 gezeigt. Bei den Knotenpunkten I bis III (siehe Abbildung 7.18) ergibt die Optimierung relativ hohe Werte (Abbildung 7.19, erste bis dritte Spur), was darauf hindeutet, dass breite Banden die Klassifizierung beeinflussen. Andere Ergebnisse liefert die Optimierung der Knotenpunkte V und VI (siehe Abbildung 7.18), bei denen relativ niedrige Werte ermittelt werden (Abbildung 7.19, fünfte und sechste Spur) und schmale Absorptionsbanden für die Klassifizierung entscheidend sind.

Die Regressionskoeffizienten der Knotenpunkte (Abbildung 7.20) zeigen, dass der Spektralbereich zwischen 1800 und 2000 cm^{-1} diese kaum beeinflusst und somit auch diese zusätzlichen Knotenpunkte primär anhand von Absorptionsbanden klassifizieren. Anhand der Regressionskoeffizienten der Knotenpunkte I, II, III und V (Abbildung 7.20, erste bis dritte

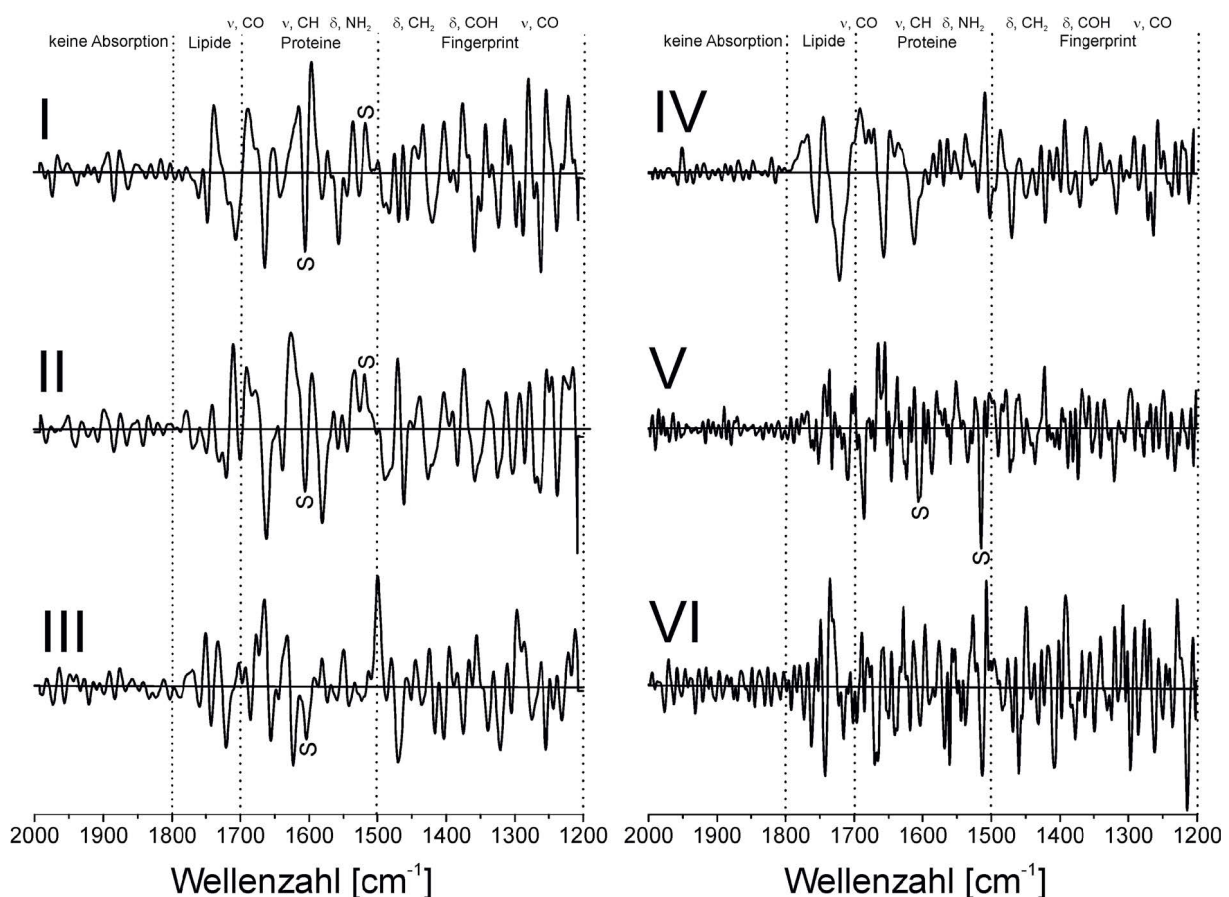


Abbildung 7.20: Regressionskoeffizienten der Knotenpunkte I bis VI des optimierten hierarchischen Klassifizierungsmodells (siehe Abbildung 7.18), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde. Sporopollenin-spezifische Banden sind mit einem S gekennzeichnet.

Spur links, sowie zweite Spur rechts) ist ersichtlich, dass für eine Identifizierung von Pollenkörnern der Ordnung *Coniferales* u.a. Sporopollenin-spezifische Banden relevant sind. Zusätzlich dazu sind sowohl Banden des Fingerprintbereichs, als auch Amidbanden, die auf Unterschiede von Proteinen hindeuten, für die Klassifizierung an diesen Knotenpunkten wichtig. Dies bestätigt die Schlussfolgerungen, die auch aus den Regressionskoeffizienten des einfachen Modells (vgl. Abbildung 7.15) gezogen wurden. Die Differenzierung von Pollenkornspektren der Gattung *Juniperus* von denen der Gattung *Betula* an Knotenpunkt IV (siehe Abbildung 7.18) basiert vor allem auf Unterschieden der Carbonylstreckschwingung von Lipiden (Abbildung 7.20, erste Spur rechts). Dies deutet darauf hin, dass *Juniperus*-Pollenkörner anhand ihrer Lipide von Pollenkörnern anderer Gattungen unterschieden werden können, was sowohl die Ergebnisse der Untersuchung von kommerziell erhältlichen Pollenproben (vgl. Abbildung 7.4), als auch die des einfachen Modells von Pollenproben aus dem Feld (vgl. Abbildung 7.15) unterstützt. Die Identifizierung von Pollenkornspektren der Gattung *Corylus* in Knotenpunkt VI (Abbildung 7.18) basiert auf Banden im gesamten Spektralbereich (Abbildung 7.20, dritte Spur rechts). Auch dies wurde bereits bei der Untersuchung von Regressionskoeffizienten des einfachen Modells (vgl. Abbildung 7.15) festgestellt.

Das erweiterte hierarchische Modell (Abbildung 7.18) wurde anschließend zur Klassifizierung des Testdatensatzes verwendet. Eine Klassifizierung auf Ebene der Ordnung (siehe Abbildung 7.21) führte dazu, dass 97% der Spektren der Ordnung *Fagales*, 89% der Spektren der Ordnung *Coniferales* und damit insgesamt 93% aller Spektren der jeweils korrekten Ordnung zugewiesen wurden. Verglichen mit den Klassifizierungsergebnissen des einfachen Modells (siehe Abbildung 7.16) werden somit elf weitere Pollenkornspektren oder zusätzlich zwei Prozent des Datensatzes korrekt klassifiziert. Durch zusätzliche Knotenpunkte können

identifiziert als			
Ordnung der getesteten Spektren		Coniferales	Fagales
Coniferales		358(89%)	45(11%)
Fagales		13(3%)	368(97%)

Abbildung 7.21: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Ordnung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach optimiertem hierarchischem Klassifizierungsmodell (siehe Abbildung 7.18), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert.

identifiziert als Gattung der getesteten Spektren	Abies	Picea	Pinus	Juniperus	Taxus	Quercus	Corylus	Alnus	Betula
Abies	43(70%)	2 (3%)	6 (10%)			6 (10%)		4 (7%)	
Picea	7 (9%)	65 (80%)	7 (9%)					1 (1%)	1 (1%)
Pinus	3 (2%)	3 (2%)	98(80%)			4 (3%)		12 (10%)	2(2%)
Juniperus				57(97%)		1 (2%)			1 (2%)
Taxus			1 (1%)		66 (83%)			4 (5%)	9 (11%)
Quercus				1 (1%)	2 (2%)	31(38%)	1 (1%)	34 (41%)	13 (16%)
Corylus						2 (2%)	108 (93%)	2 (2%)	4 (3%)
Alnus	2 (3%)		1 (1%)	1 (1%)		1 (1%)	8 (10%)	67 (84%)	
Betula	2 (2%)	1 (1%)	3 (3%)			8 (8%)	5 (5%)	53 (51%)	31 (30%)

Abbildung 7.22: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Gattung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach optimiertem hierarchischem Klassifizierungsmodell (siehe Abbildung 7.18), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert.

also Varianzen hervorgehoben werden, durch die Fehlklassifizierungen von Proben verschiedener Ordnungen ausgeglichen werden können.

Anschließend wurde das Modell auch für die Klassifizierung auf Ebene der Gattung verwendet. Das Ergebnis ist in Abbildung 7.22 gezeigt. Da keine zusätzlichen Knotenpunkte für die Unterscheidung der Gattungen *Alnus*, *Quercus* und *Betula* hinzugefügt werden konnten, scheitert auch in diesem Modell ihre Unterscheidung. Wendet man allerdings das optimierte hierarchische Modell auf die Spektren der anderen Gattungen an, werden insgesamt 84% dieser taxonomischen Gruppen (Abbildung 7.22) korrekt klassifiziert. Durch die zusätzlichen Knotenpunkte kann insbesondere die fehlerhafte Zuordnung von Spektren der Gattungen *Pinus*, *Abies* und *Corylus* verringert werden (vgl. mit Abbildung 7.17). Obwohl in den Spektren charakteristische Absorptionsbanden mit störenden Streuartefakten überlagern, können also die meisten Spektren einzelner Pollenkörner der Gattungen *Abies*, *Picea*, *Pinus*, *Juniperus*, *Taxus* und *Corylus* anhand dieses Modells unterschieden werden. Dabei kann die Fehlklassifizierung einzelner Spektren, die aus dem Einfluss von Streubeiträgen resultiert, durch zusätzliche Knotenpunkte im Modell ausgeglichen werden. Diese sind auf die Unterschiede zwischen spezifischen Klassen fokussiert.

7.3 Schlussfolgerungen

Allgemein kann mit FTIR-Spektroskopie einzelner Pollenkörner im Transmissionsmodus ein spezifischer, taxonomisch relevanter Fingerabdruck der Pollenprobe erhalten werden. Durch ein geeignetes hierarchisches PLSR-Klassifizierungsmodell^[164] basierend auf einer Optimierung der SG-Datenpunkte bei der zweiten Ableitung und einer EMSC-Korrektur der Spektren^[163] kann in vielen Fällen eine erfolgreiche taxonomische Klassifizierung dieser Fingerabdrücke erreicht werden. Durch die unterschiedliche Anzahl an SG-Datenpunkten beim Bilden der zweiten Ableitung werden die verschiedenen spektralen Beiträge in diesen hierarchischen Modellen unterschiedlich gewichtet. Wird eine relativ hohe Anzahl an SG-Datenpunkten verwendet, sind breite Banden, wie beispielsweise Amidbanden, für die Klassifizierung ausschlaggebend. Da es sich bei oszillierenden Streuartefakten ebenfalls um relativ breite Banden handelt, kann ihr Einfluss zusätzlich verringert werden, indem eine relativ geringe Anzahl an SG-Datenpunkten für die Klassifizierung verwendet wird. Sind charakteristische Absorptionsbanden einzelner Spektren mit Streuartefakten überlagert, ist es möglich, zusätzliche Knotenpunkte im hierarchischen Klassifizierungsmodell einzufügen, die auf die Differenzierung spezifischer, taxonomischer Gruppen fokussiert sind. Auf diese Weise können verschiedene charakteristische Varianzen für die taxonomische Klassifizierung in das Modell integriert werden, was die Wahrscheinlichkeit einer zutreffenden Klassifizierung erhöht. Die Unterscheidung von Pollenproben verschiedener Gattungen ist allerdings schwierig, wenn diese ähnliche Eigenschaften aufweisen. Pollenkörner der Ordnung *Fagales* haben sowohl sehr ähnliche Absorptionsspektren^[17], die auf eine ähnliche chemische Zusammensetzungen schließen lässt, als auch eine vergleichbare Größe, Form und Morphologie, die zu ähnlichen Streuartefakten in den FTIR-Spektren führt. In diesem Fall sind Streuartefakte in den Spektren ausschlaggebend dafür, dass zwischen den Absorptionsspektren einzelner Pollenkörner verschiedener Gattungen häufig nicht unterschieden werden kann. Um auch diese ähnlichen Pollenproben durch FTIR-Spektroskopie einzelner Pollenkörner differenzieren zu können, ist es notwendig die Beschaffenheit der störenden Streuartefakte weiter zu untersuchen und Algorithmen zu entwickeln, die klassifizierungsrelevante Absorptionsbanden extrahieren können. Eine Weiterentwicklung in diesem Bereich wurde erst kürzlich veröffentlicht^[27] und könnte auch in Kombination mit den hier angewendeten Algorithmen eine erfolgreiche Klassifizierung einer größeren Anzahl an taxonomischen Gruppen ermöglichen.

8. Klassifizierung CLRV-infizierter Birkenpollen

Beim *cherry leaf roll virus* (CLRV) handelt es sich um ein Pflanzenvirus, das verschiedene Wirtspflanzen befallen kann und in Ländern auf der ganzen Welt auftritt.^{[166]-[168]} Erstmals dokumentiert wurde es 1933 in Walnussbäumen (*Juglans regia*),^[169] anschließend aber auch in anderen Bäumen, wie z.B. der Süßkirsche (*Prunus avium*)^[170] und der Birke (*Betula pendula*)^[171] nachgewiesen. Eine Infektion der jeweiligen Pflanzen äußert sich vor allem durch chlorotische Ringflecken und Linienmuster auf den Blättern, sowie eine Blattscheckung und -kräuselung,^[171] was auf eine Veränderung der Ultrastruktur der Blattzellen zurückgeführt werden kann^[172]. Da CLRV zwei RNA-Strängen mit einer Gesamtlänge von ca. 7900 bzw. 6400 Basen enthält,^{[173],[174]} wurde es bereits in den 70er Jahren des letzten Jahrhunderts als Virus der Gattung Nepovirus klassifiziert.^[175] Bei detaillierten Analysen der RNA-Sequenzen stellte sich sowohl heraus, dass diese abhängig davon, aus welcher Pflanze sie isoliert wurden, unterschiedlich sein können,^[166] als auch, dass Abschnitte existieren, die jeweils identisch auf beiden RNA-Strängen vorliegen.^[176] Da CLRV nicht nur über die Blätter, sondern auch über Samen^[177] und Pollen übertragen werden kann,^[178] ist eine Identifizierung des Virus auch anhand der Untersuchung von Pollenproben möglich. Aktuell wird CLRV mit einer *immunocapture reverse transcription-polymerase chain reaction* (IC-RT-PCR) nachgewiesen.^[179] Dabei wird die RNA des Virus durch spezifische Antikörper isoliert und durch eine reverse Transkriptase in eine cDNA umgeschrieben, welche durch die PCR anschließend amplifiziert wird. Da diese Methode allerdings sehr zeitaufwendig ist, ist es sinnvoll, nach Alternativen für die Identifizierung von CLRV-infizierten Proben zu suchen. Es wurde gezeigt, dass eine Virusinfektion eine Immunantwort der Pflanze hervorruft und sich damit auf die chemische Zusammensetzung von Pollenproben auswirkt.^[180] Deshalb ist es wahrscheinlich, dass es möglich ist, spektroskopische und spektrometrische Methoden für die Identifizierung von CLRV-infizierten Birkenpollenproben einzusetzen. Im Rahmen dieser Arbeit wurden Pollenproben gesammelt, die anschließend mit IC-RT-PCR auf eine CLRV-Infektion getestet und mit *surface-enhanced Raman scattering* (SERS) (vgl. Kapitel 4), *matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS) (vgl. Kapitel 5) und Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie) von Kaliumbromidpresslingen (vgl. [16]) untersucht wurden. Anschließend wurde überprüft, ob die Daten der CLRV-infizierten Proben Ähnlichkeiten aufwiesen und eine Infektion daher identifiziert werden kann. Es wurden dabei auch Pollenproben eines Individuums analysiert, bei dem zwar keine CLRV-Infektion, aber eine Infektion mit dem Tabakmosaikvirus (TMV)

nachgewiesen werden konnte. Dadurch kann untersucht werden, ob die ggf. auftretenden Unterschiede der spektralen Fingerabdrücke spezifisch für die CLRV-Infektion sein könnten. Die Produktion von Birkenpollen unterliegt einem Zweijahreszyklus, bei dem sich sogenannte Mastjahre, wie das Jahr 2012, mit Jahren, in denen eine bedeutend geringere Anzahl an Pollen produziert werden, abwechseln. Um zu testen, ob sich die Intensität der Pollenproduktion auf die Untersuchung CLRV-infizierter Pollenproben auswirkt, wurden die Pollenproben der Jahre 2011 (Abschnitt 8.1) und 2012 (Abschnitt 8.2) separat untersucht.

8.1 Untersuchung von Birkenpollenproben aus dem Jahr 2011

Es wurden im Jahr 2011 Birkenpollenproben gesammelt und mit IC-RT-PCR von Kooperationspartnern (Arbeitsgruppe von Prof. Büttner der Lebenswissenschaftlichen Fakultät an der Humboldt-Universität zu Berlin) auf eine CLRV-Infektion, bzw. auf eine TMV-Infektion getestet. Es wurden die SERS-Untersuchung wässriger Pollenextrakte (Abschnitt 8.1.1), MALDI-TOF MS von Säureextrakten (Abschnitt 8.1.2), sowie die FTIR-Untersuchung von Kaliumbromidpresslingen (Abschnitt 8.1.3) durchgeführt. Die Daten, die Unterschiede zwischen CLRV-infizierten und gesunden Birkenpollenproben aufzeigten, wurden anschließend auch mit *Consensus Principal Component Analysis* (CPCA) (Abschnitt 8.1.4) in einer Auswertung, die einen größeren Teil der komplexen Pollenchemie abbildet, kombiniert.

8.1.1 Klassifizierung CLRV-infizierter Birkenpollenproben mit SERS

Im Rahmen dieser Arbeit wurde gezeigt, dass durch SERS von wässrigen Pollenextrakten die inneren, zellulären Bestandteile von Pollenkörnern untersucht und klassifiziert werden können (siehe Abschnitt 4). Um zu untersuchen, ob anhand des SERS-Fingerabdrucks der Extrakte der jeweiligen Pollenproben zwischen CLRV-infizierten und gesunden Pollenproben differenziert werden kann, wurden aus den vorbearbeiteten Pollendaten der Hälfte der Proben ein künstliches neuronales Netz (*artificial neural network*, ANN) trainiert und dieses mit unabhängigen, vorbearbeiteten Daten der anderen Hälfte der Proben getestet. Abbildung 8.1 zeigt die Ergebnisse der Klassifizierung. Durch SERS des wässrigen Pollenextraktes kann demnach keine Unterscheidung zwischen CLRV-infizierten und gesunden Birken erfolgen. Die SERS-Spektren der wässrigen Pollenextrakte weisen also keine signifikanten und spezifischen Merkmale auf, die für die Identifizierung einer CLRV-Infektion verwendet werden

Ergebnisse der IC-RT-PCR	klassifiziert als	
	positiv	negativ
CLRV positiv	9 (35%)	15 (65%)
CLRV negativ	6 (36%)	8 (57%)
CLRV negativ, TMV positiv	4 (67%)	2 (33%)

Abbildung 8.1: Ergebnis der ANN-Klassifizierung von 46 SERS-Datensätzen von Extrakten aus Birkenpollen aus dem Jahr 2011. Es wurden jeweils die Zuordnungen von 1000 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

können. Ein Grund dafür ist sicherlich, dass die chemische Zusammensetzung der Pollenkörner und damit auch die der inneren, wasserlöslichen Pollenbestandteile auch durch andere Umweltfaktoren beeinflusst werden. Durch FTIR-Spektroskopie wurde beispielsweise bereits belegt, dass Umweltparameter den Gehalt an in den Pollenkörnern enthaltenen Lipiden signifikant beeinflussen.^[17]

8.1.2 Klassifizierung CLRV-infizierter Birkenpollenproben mit MALDI-TOF MS

Es wurde bereits gezeigt, dass durch MALDI-TOF MS-Untersuchungen von Säureextrakten Pollenproben taxonomisch klassifiziert werden können (siehe Abschnitt 5). Die spezifischen, taxonomischen Fingerabdrücke können dabei vor allem auf Glykoproteine der Pollenproben zurückgeführt werden.^[134] Durch MALDI-TOF MS von Birkenpollenproben soll nun überprüft werden, ob anhand dieser, v.a. durch Glykoproteine beeinflusster Fingerabdrücke eine Differenzierung in CLRV-positiv getestete und gesunde Individuen erfolgen kann. Vergleicht man jeweils repräsentative Spektren der Proben (Abbildung 8.2), so ist erkennbar, dass beide die für Birkenpollen charakteristischen Peaks aufweisen (vgl. Abbildung 5.3). Eine detailliertere Betrachtung zeigt allerdings, dass Unterschiede zwischen CLRV-positiv getesteten und gesunden Proben auftreten. So weisen die Spektren CLRV-negativ getesteter Pollenproben (Abbildung 8.2, schwarzes Spektrum), Peaks bei m/z 6819, 6832 und 7148 auf, während der Peak bei 7116 vergleichsweise wenig intensiv ist. Im Gegensatz dazu sind bei CLRV-positiv getesteten Proben (Abbildung 8.2, rotes Spektrum), Peaks bei m/z 6791 und 6805 erkennbar und der Peak bei m/z 7118 ist bedeutend intensiver ausgeprägt.

Die MALDI-TOF MS-Daten wurden anschließend durch eine Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) analysiert. Die Auftragung der ersten beiden Hauptkomponenten (Abbildung 8.3a), zeigt, dass die zweite Hauptkomponente für die Differenzie-

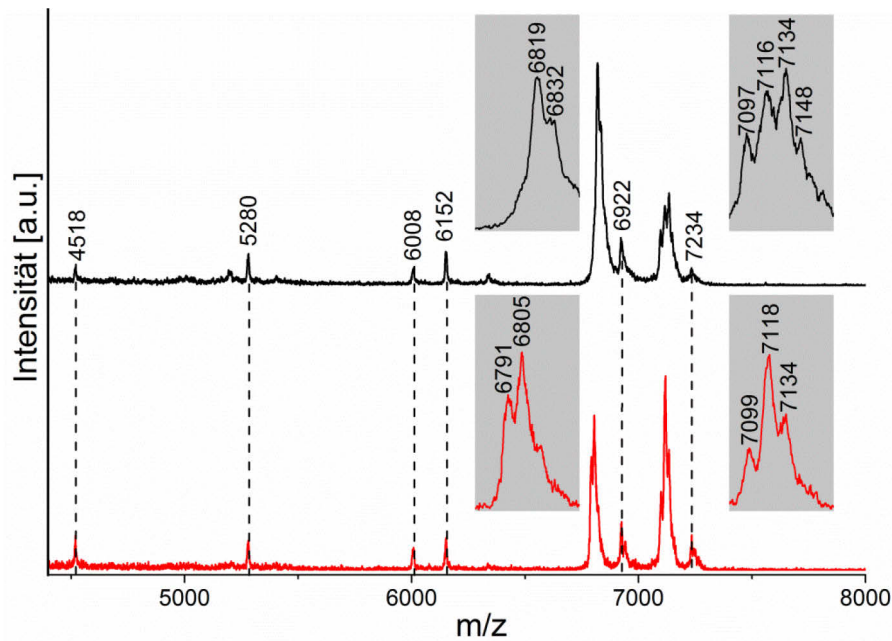


Abbildung 8.2: Vergleich repräsentativer MALDI-TOF MS-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2011. Es ist jeweils der Massenbereich zwischen m/z 4400 und 8000, sowie die vergrößerten Massenbereiche zwischen m/z 6760 und 6860 und m/z 7080 und 7180 (in grau) gezeigt.

zung zwischen CLRV-positiv getesteten und gesunden Pollenproben verwendet werden kann, während die erste Hauptkomponente anscheinend nicht durch die CLRV-Infektion beeinflusst wird. Die erste Hauptkomponente, die eine Varianz von fast 40% repräsentiert, deutet somit auf zusätzliche, intensive Varianzen hin, die unabhängig von der CLRV-Infektion im Datensatz auftreten. Ersetzt man die Scores dieser Hauptkomponente durch die der dritten (Abbildung 8.3b), ist eine Unterscheidung fast aller Proben in virusbefallene und gesunde Pollenproben möglich. Die Spektren TMV-positiver Birkenpollen weisen dabei ähnliche Scores auf wie die CLRV-positiv getesteten Pollenproben. Die Ergebnisse lassen also vermuten, dass mit MALDI-TOF MS Veränderungen der Pollenchemie detektiert werden könnten, die generell durch einen Virus, aber nicht unbedingt durch CLRV verursacht werden. Die Loadings der zweiten und dritten Hauptkomponente (Abbildung 8.3c, zweite und dritte Spur), die eine Varianz von ca. 30% repräsentieren, zeigen, dass die Peaks bei m/z 6791, 6812 und 7117 entscheidend zu dieser Unterscheidung beitragen. Diese konnten bereits beim Betrachten der Spektren (siehe Abbildung 8.2) als für die Differenzierung signifikant identifiziert werden.

Anhand der MALDI-TOF MS-Untersuchung von Birkenpollenproben ist also eine Identifizierung von virusinfizierten Proben möglich. Da diese vor allem auf einer

Untersuchung der Glykoproteine basiert und es sich bei den Allergenen der Birkenpollen ebenfalls um Glykoproteine handelt, deuten diese Ergebnisse darauf hin, dass auch Allergene und das allergische Potential von Birkenpollen durch eine Virusinfektion verändert werden könnten. Ein derartiger Einfluss von Pflanzenviren auf Allergene und damit die allergische Reaktion des Menschen wurde bereits anhand von virusinfizierten Graspollen anhand eines Hauttestes festgestellt,^[29] was diese Vermutung unterstützt.

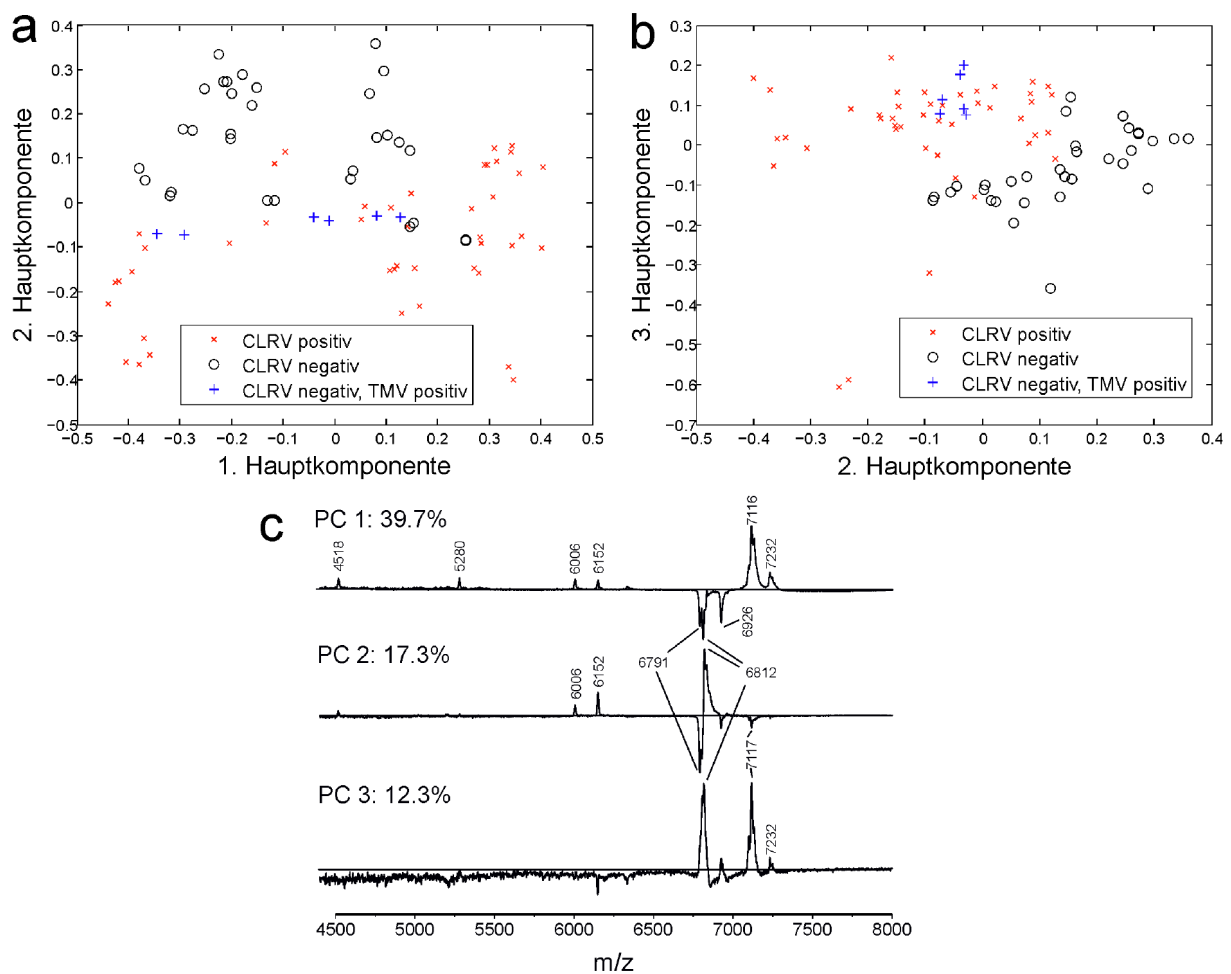


Abbildung 8.3: Ergebnis der PCA von 82 MALDI-TOF MS-Spektren zur Unterscheidung CLRV-positiv und CLRV-negativ getesteter Birkenpollenproben aus dem Jahr 2011. Es sind die Scores der ersten und zweiten Hauptkomponente (a), die der zweiten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der ersten drei Hauptkomponenten (c) gezeigt.

8.1.3 Klassifizierung von Kaliumbromidpresslingen CLRV-infizierter Birkenpollenproben mit FTIR-Spektroskopie

Durch die FTIR-Spektroskopie von Kaliumbromidpresslingen werden spezifische Pollendaten erhalten, die für eine taxonomische Klassifizierung verwendet werden können.^[16] Die Banden in den FTIR-Spektren können dabei zumeist Lipiden, Proteinen, Kohlenhydraten oder Sporopollenin, dem Makromolekül der Pollenaußenhülle, zugeordnet werden. Da mit den bisher verwendeten Methoden vor allem die inneren, zellulären Bestandteile (siehe Abschnitt 8.1.1), bzw. Glykoproteine (siehe Abschnitt 8.1.2) untersucht wurden, werden mit FTIR-Untersuchung von Kaliumbromidpresslingen komplementäre Informationen erhalten, die auch Informationen der äußeren Pollenbestandteile enthalten. Vergleicht man repräsentative Spektren von CLRV-positiv und CLRV-negativ getesteten Birkenpollenproben (Abbildung 8.4), so ist erkennbar, dass beide die charakteristischen Banden von Pollenproben enthalten (Zuordnung in Tabelle 8.1) und keine offensichtlichen Unterschiede zwischen den Spektren auftreten. Um zu überprüfen, ob eine Differenzierung anhand von kleinen, nicht mit bloßem Auge erkennbaren, Unterschieden in den Spektren erfolgen kann, wurde anschließend eine PCA durchgeführt. Ähnlich wie bei der Untersuchung mit MALDI-TOF MS (vgl. Abschnitt 8.1.2) sind jedoch intensive Varianzen im Datensatz vorhanden, die nicht auf Differenzen zwischen virusbefallenen und gesunden Pollenproben zurückgeführt werden können. Dies

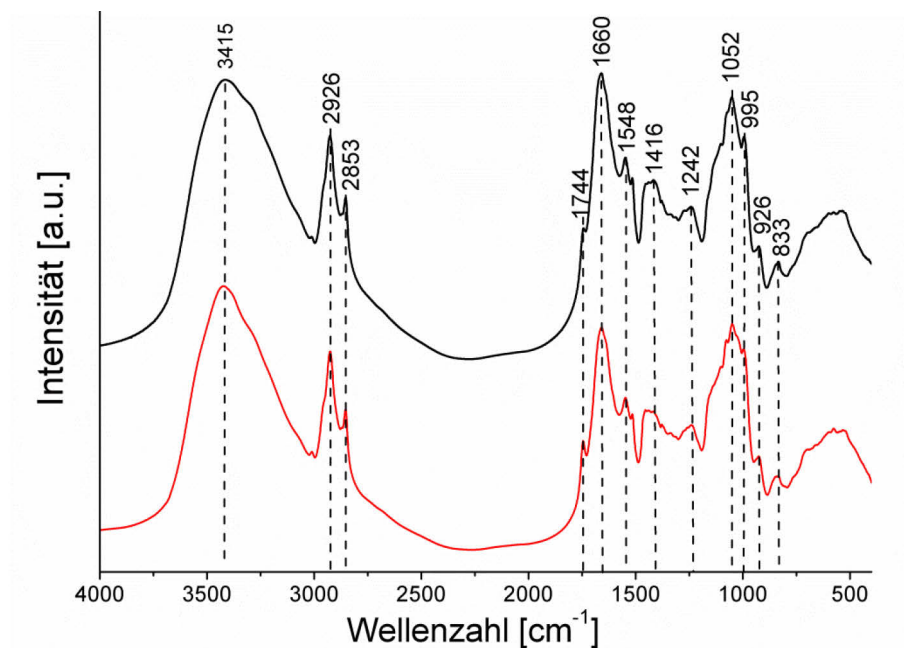


Abbildung 8.4: Vergleich repräsentativer FTIR-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2011.

Tabelle 8.1: Charakteristische Absorptionsbanden der FTIR-Spektren von Birkenpollenproben aus dem Jahr 2011 und deren Zuordnung zu Referenzspektren

Wellenzahl [cm^{-1}]	Zuordnung ^a
3415	v, OH
2926	v, CH ₃ asym. ^c
2853	v, CH ₃ sym. ^c
1744	v, C=O, Lipide ^b
1660	v, C=O, Proteine, Amid I ^b
1548	v, C-N, δ , N-H, Proteine, Amid II ^b
1416	δ , CH ₂ , Lipide ^b
1242	v, C-O, Lipide ^b
1052	v, C-O-C, Kohlenhydrate ^{b,c}
995	v, Kohlenhydrate ^{b,c}
926	Ringatmung, Kohlenhydrate ^{b,c}
833	Ringatmung, Sporopollenin ^b

^a Abkürzungen: v, Streckschwingung δ , Deformationsschwingung
sym, symmetrisch asym, asymmetrisch

^b nach [16] und [17]

^c nach [181]

wird anhand der Scores der ersten Hauptkomponente (Abbildung 8.5a), die eine Varianz von 42,5% repräsentiert (Abbildung 8.5b), deutlich. Anhand der zweiten Hauptkomponente, die eine Varianz von 16,6 % repräsentiert (Abbildung 8.5b), können jedoch virusbefallene Proben mit eher positiven Scores von gesunden Pollenproben mit eher negativen Scores (Abbildung 8.5a) differenziert werden. Die Loadings der zweiten Hauptkomponente (Abbildung 8.5b, zweite Spur) zeigen, dass verschiedene Banden zu dieser Differenzierung beitragen. Diese können größtenteils sowohl Kohlenhydraten (um 1000 und 2900 cm^{-1}) als auch Lipiden und Proteinen (zwischen 1500 und 1750 cm^{-1}) zugeordnet werden (siehe Tabelle 8.1). Die Ergebnisse der FTIR-Untersuchung könnten somit ebenso wie die Untersuchung mit MALDI-TOF MS (vgl. Abschnitt 8.1.2) auf Differenzen in der Glykoproteinzusammensetzung zwischen virusbefallenen und gesunden Pollenproben hindeuten. Interessant ist, dass die Bande bei ca. 1385 cm^{-1} einen beträchtlichen Einfluss auf die zweite Hauptkomponente (Abbildung 8.5b, zweite Spur) und damit auf die Klassifizierung von virusbefallenen Pollenproben mit FTIR-Spektroskopie von Kaliumbromidpresslingen ausübt. Diese Bande kann mit bloßem Auge in den Spektren nur erahnt werden (siehe Abbildung 8.4) und ist vermutlich auf eine Deformationsschwingung von Kohlenhydraten zurückzuführen. In den Loadings der zweiten Hauptkomponente sind keine Banden von Sporopollenin erkennbar.

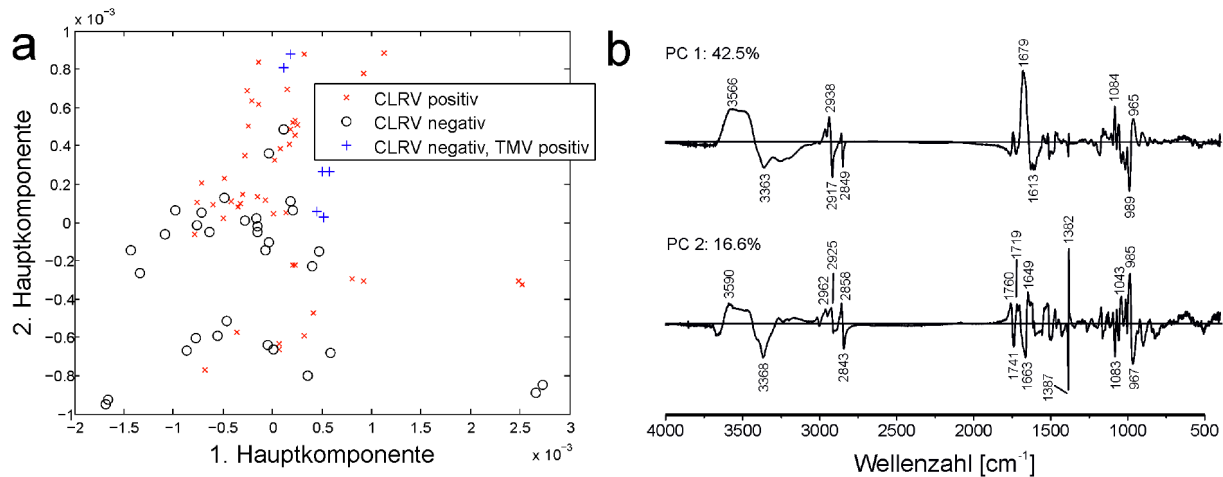


Abbildung 8.5: Ergebnis der PCA von 82 FTIR-Spektren zur Unterscheidung von Kaliumbromidpresslingen von CLRV-positiv und CLRV-negativ getesteten Birkenpollenproben aus dem Jahr 2011. Es sind die Scores der ersten und zweiten Hauptkomponente (a), sowie die Loadings und Varianzen der ersten beiden Hauptkomponenten (c) gezeigt.

Dies deutet darauf hin, dass die Außenhülle durch die Virusinfektion weitestgehend unverändert bleibt. Auch in der PCA der FTIR-Spektren weisen die TMV-infizierten Pollenproben (Abbildung 8.5a), ähnliche Scores wie die CLRV-infizierten Proben auf. Auch die FTIR-Untersuchungen deuten also auf Unterschiede der Pollenkörner hin, die durch einen Virus aber nicht zwangsläufig durch CLRV verursacht werden.

Mit FTIR-Spektroskopie von Kaliumbromidpresslingen kann also eine Unterscheidung vieler Proben in virusbefallene und gesunde Pollenproben erfolgen. Da die spektralen Unterschiede, die diese Separierung bewirken nicht auf Unterschiede in der Sporopolleninzusammensetzung, sondern auf solche von Kohlenhydraten und Proteinen, bzw. Lipiden hindeuten, ist es wahrscheinlich, dass eine Virusinfektion primär Bestandteile verändert, die sich in den zellulären, inneren Bestandteilen von Pollenkörnern befinden. Dies unterstützt die Ergebnisse von MALDI-TOF MS und könnte ebenfalls auf eine virusbedingte Änderung des allergenen Potentials von Pollenkörnern hindeuten.^[29]

8.1.4 Kombination verschiedener analytischer Daten zur Klassifizierung CLRV-infizierter Birkenpollenproben

In den vorigen Abschnitten (siehe 8.1.2 und 8.1.3) wurde gezeigt, dass sowohl mit MALDI-TOF MS, als auch mit FTIR-Spektroskopie von Kaliumbromidpresslingen, virusspezifische Varianzen in der Zusammensetzung der Pollenkörner detektiert werden können. Um in der Klassifizierung einen möglichst großen Teil der komplexen chemischen Zusammensetzung

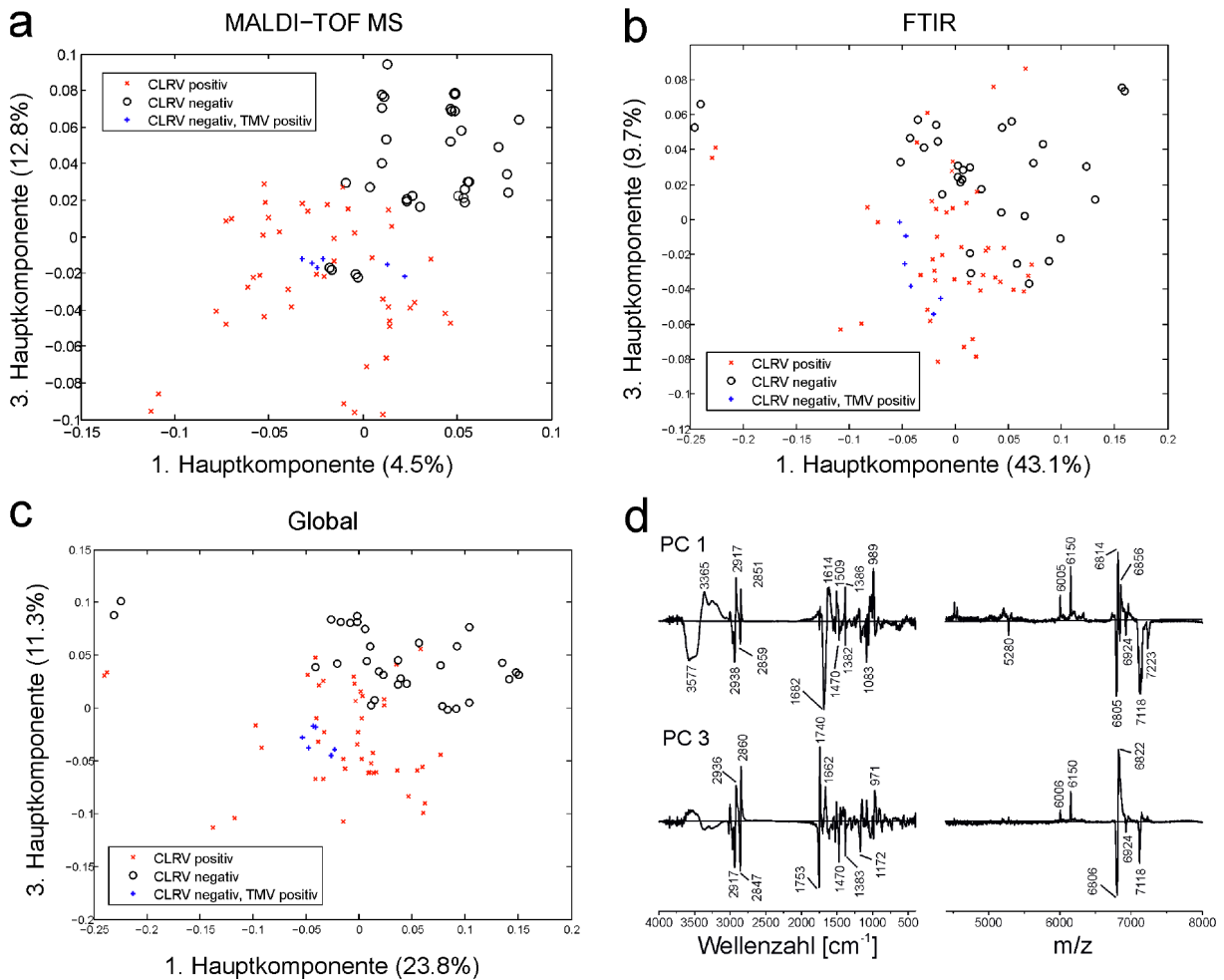


Abbildung 8.6: Ergebnis der CPCA von MALDI-TOF MS- und FTIR-Daten zur Unterscheidung von 82 Datensätzen in CLRV-positiv und CLRV-negativ getestete Birkenpollenproben aus dem Jahr 2011. Die Scores und Varianzen der ersten und dritten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), sowie die Loadings der ersten und dritten CPCA-Komponente (d) sind gezeigt.

von Pollenkörnern abzubilden, wurden anschließend die Daten dieser Methoden mit einer CPCA kombiniert. Abbildung 8.6 zeigt die Ergebnisse dieser Auswertung. Anhand der ersten und vor allem der dritten Hauptkomponente der einzelnen Blöcke (Abbildung 8.6a und b) und des globalen Blocks (Abbildung 8.6c) können die Daten von Pollenproben virusinfizierter Bäume von denen gesunder Bäume unterschieden werden. Da die dritte Hauptkomponente für beide Blöcke mit 12.8% (Abbildung 8.6a), bzw. 9.7% (Abbildung 8.6b) eine ähnliche erklärte Varianz aufweist, lässt sich schließen, dass beide Blöcke zu der Differenzierung beitragen. Der Vergleich der Scores des globalen Blocks (Abbildung 8.6c) mit denen der PCA der einzelnen Blöcke (vgl. Abbildung 8.3, bzw. Abbildung 8.5), zeigt, dass eine eindeutigere Unterscheidung zwischen den Pollenproben virusinfizierter und gesunder Bäume möglich ist. Ähnlich wie bei der CPCA von SERS- und MALDI-TOF MS-Daten für die taxonomische

Klassifizierung von Pollenproben (vgl. Abschnitt 6) führt somit die Kombination von komplementären Daten verschiedener analytischer Methoden zu einer robusteren Klassifizierung. Die Loadings der dritten CPCA-Hauptkomponente des FTIR-Blocks (Abbildung 8.6d, zweite Spur links) weisen charakteristische Banden von Lipiden (ca. 1745 und 1470 cm^{-1}), Peptiden (ca. 1650 cm^{-1}) und Kohlenhydraten (ca. 1000 und 1380 cm^{-1}) auf, während in denen des MALDI-TOF MS-Blocks (Abbildung 8.6d, zweite Spur rechts) charakteristische Massen, z.B. bei m/z 6808 und 7118 identifiziert werden können. Der Vergleich der Loadings mit denen der PCA der einzelnen Blöcke (Abbildung 8.3c und Abbildung 8.5b) zeigt, dass die Identifizierung virusinfizierter Pollenproben größtenteils auf Varianzen basiert, die bereits in den einzelnen Blöcken als virusspezifisch identifiziert wurden. Dass der Einfluss der lipidspezifischen FTIR-Banden bei ca. 1745 und 1470 cm^{-1} (Abbildung 8.6d, zweite Spur links) bedeutend größer ist als bei der PCA, die nur auf der Varianz der FTIR-Daten basiert (siehe Abbildung 8.5), legt die Vermutung nahe, dass neben Glykoproteinen auch Lipidbestandteile von Pollenkörnern durch die Virusinfektion verändert werden. Sporopollenin-spezifische FTIR-Banden treten auch in den FTIR-Loadings der CPCA-Hauptkomponenten (Abbildung 8.6) nicht auf, was darauf hindeutet, dass die Außenhülle von Pollenkörnern durch die Virusinfektion nicht signifikant beeinflusst wird. Die Untersuchung von Proben aus dem Jahr 2011 zeigt also, dass eine Virusinfektion von Birken anhand von MALDI-TOF MS- (Abschnitt 8.1.2) und FTIR-Untersuchungen (Abschnitt 8.1.3) von Pollenkörnern in Kombination mit multivariater Auswertung detektiert werden kann. Die Ergebnisse deuten darauf hin, dass sich Glykoprotein- und Lipidzusammensetzung der inneren Pollenbestandteile bei virusbefallenen Proben unterscheiden. Deshalb könnte auch das allergene Potential von Pollen durch eine Virusinfektion verändert werden.^[29]

8.2 Untersuchung von Birkenpollenproben aus dem Jahr 2012

Um zu überprüfen, ob sich die Intensität der Pollenproduktion auf die Klassifizierung von Pollenproben CLRV-infizierter und gesunder Birken auswirkt, wurden anschließend auch Pollenproben aus dem Mastjahr 2012 untersucht. Äquivalent zur Untersuchung der Proben aus dem Jahr 2011 (Abschnitt 8.1) wurden die SERS-Untersuchung wässriger Pollenextrakte (Abschnitt 8.2.1), MALDI-TOF MS von Säureextrakten (Abschnitt 8.2.2), sowie die FTIR-Untersuchung von Kaliumbromidpresslingen (Abschnitt 8.2.3) durchgeführt.

8.2.1 Klassifizierung CLRV-infizierter Birkenpollenproben mit SERS

Anhand der SERS-Fingerabdrücke von Pollenextrakten CLRV-infizierter und gesunder Pollenproben aus dem Jahr 2012 soll untersucht werden, ob eine CLRV-Infektion dieser Proben mit SERS identifiziert werden kann. Deshalb wurde aus den Pollendaten der einen Hälfte der Proben ein ANN trainiert und dieses mit unabhängigen SERS-Daten von Extrakten der anderen Hälfte der Proben getestet. Abbildung 8.7 zeigt die Ergebnisse der Klassifizierung durch das ANN. Es werden zwar 91% der CLRV-positiv getesteten Proben korrekt klassifiziert, jedoch werden auch über 70% der CLRV-negativ getesteten Proben als CLRV-positiv eingeordnet. Ähnlich wie die Proben des Jahres 2011 können somit auch Birkenpollenproben aus dem Mastjahr 2012 nicht anhand der SERS-Untersuchung von wässrigen Extrakten in CLRV-positiv und CLRV-negativ differenziert werden. Auch hier könnte die Ursache für diese inkorrekte Klassifizierung darin liegen, dass zusätzliche nicht CLRV-relevante Parameter signifikant die SERS-Daten beeinflussen und CLRV-spezifische Varianzen nicht ausreichend intensiv in diesen repräsentiert sind.

Ergebnisse der IC-RT-PCR	klassifiziert als	
	positiv	negativ
CLRV positiv	29 (91%)	3 (9%)
CLRV negativ	37 (74%)	13 (26%)
CLRV negativ, TMV positiv	2 (100%)	

Abbildung 8.7: Ergebnis der ANN-Klassifizierung von 84 Datensätzen von Extrakten aus Birkenpollen aus dem Jahr 2012. Es wurden jeweils die Zuordnungen von 1000 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt.

8.2.2 Klassifizierung CLRV infizierter Birkenpollenproben mit MALDI-TOF MS

Im Gegensatz zur Untersuchung von wässrigen Extrakten mit SERS konnte anhand der MALDI-TOF MS-Analyse von Säureextrakten der Proben aus dem 2011 zwischen CLRV-positiven und -negativen Pollenproben differenziert werden (siehe Abschnitt 8.1.2). Um zu testen, ob diese Unterscheidung auch anhand von Proben, die in einem Mastjahr gesammelt wurden, erfolgreich durchgeführt werden kann, wurden Proben des Jahres 2012 auf gleiche Weise analysiert. Abbildung 8.8 zeigt den Vergleich von MALDI-TOF MS-Spektren CLRV-positiv (rot) und CLRV-negativ (schwarz) getesteter Pollenproben. Allgemein weisen beide Spektren Peaks auf, die als charakteristisch für Birkenpollenproben identifiziert wurden (vgl.

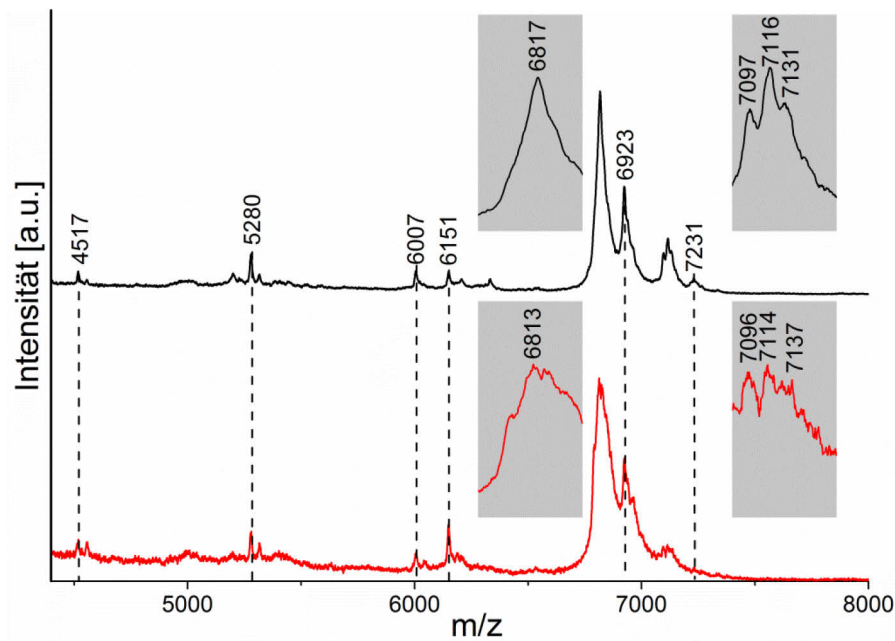


Abbildung 8.8: Vergleich von MALDI-TOF MS-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2012. Es ist jeweils der Massenbereich zwischen m/z 4400 und 8000, sowie die vergrößerten Massenbereiche zwischen m/z 6760 und 6860 und m/z 7080 und 7180 (in grau) gezeigt.

Abbildung 5.3). In Abschnitt 8.1.2 wurde in den Spektren CLRV-positiver Pollenproben Massen z.B. bei m/z 6791 und 6805 identifiziert, die sich signifikant von denen CLRV-negativ getesteter Pollenproben unterschieden, da diese charakteristische Massen z.B. bei m/z 6819 und 6832 aufwiesen (siehe Abbildung 8.2). Diese Unterschiede finden sich allerdings in den Spektren der Birkenpollenproben des Mastjahres 2012 nicht wieder (Abbildung 8.8). Dies deutet darauf hin, dass Proben, die in Jahren unterschiedlicher Pollenproduktionsintensität gesammelt wurden, signifikante Unterschiede aufweisen. Um zu überprüfen, ob trotz der jeweils unterschiedlichen Fingerabdrücke eine Differenzierung zwischen CLRV-positiven und CLRV-negativen Pollenproben anhand von MALDI-TOF MS-Untersuchungen möglich ist, wurden die Spektren anschließend mit PCA ausgewertet. Die Scores der ersten drei Hauptkomponenten (Abbildung 8.9a und b) zeigen keine signifikanten Unterschiede zwischen CLRV-positiven und CLRV-negativen Pollenproben auf. Vergleicht man die Loadings der Hauptkomponenten (Abbildung 8.9c) mit denen der PCA von Pollenproben aus dem Jahr 2011 (siehe Abbildung 8.3c) ist erkennbar, dass die Varianzen des MALDI-TOF MS-Datensatzes prinzipiell sehr unterschiedlich sind und die charakteristischen Massen CLRV-infizierter Pollenproben, nicht in den Spektren auftreten. Dies deutet darauf hin, dass eine Virusinfektion von Birken mit MALDI-TOF MS nur dann nachweisbar ist, wenn

Pollenproben eines Jahres untersucht werden, in dem eine geringe Menge an Pollen produziert werden.

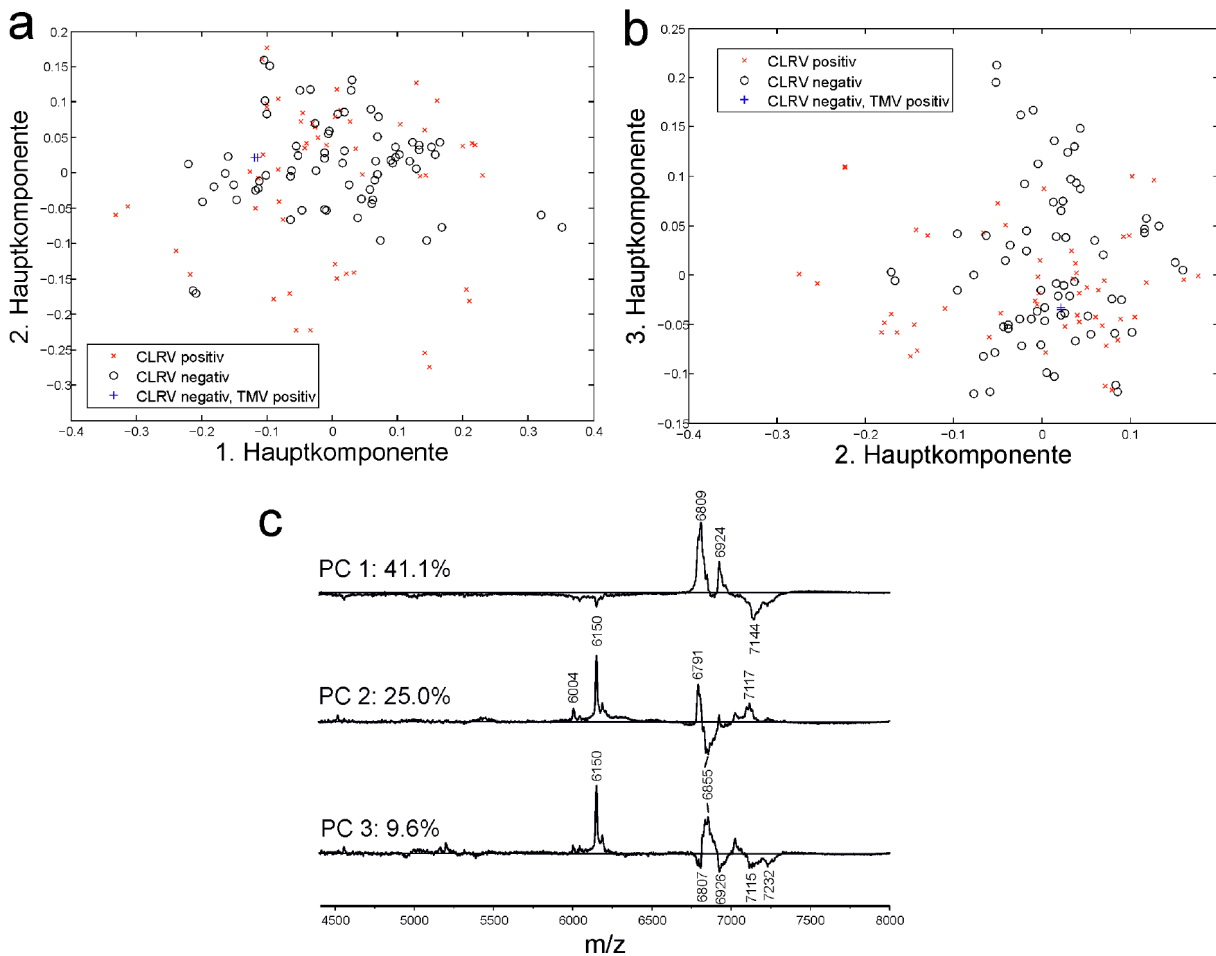


Abbildung 8.9: Ergebnis der PCA von 124 MALDI-TOF MS-Spektren zur Unterscheidung CLRV-positiv und CLRV-negativ getesteter Birkenpollenproben aus dem Jahr 2012. Es sind die Scores der ersten und zweiten Hauptkomponente (a), die der zweiten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der ersten drei Hauptkomponenten (c) gezeigt.

8.2.3 Klassifizierung von Kaliumbromidpresslingen CLRV-infizierter Birkenpollenproben mit FTIR-Spektroskopie

Durch die FTIR-Spektroskopie von Kaliumbromidpresslingen wird ein anderer Teil der Pollenchemie untersucht, als durch MALDI-TOF MS von Säureextrakten. Deshalb könnte es trotz der fehlerhaften Klassifizierung durch MALDI-TOF MS (siehe Abschnitt 8.2.2) möglich sein, eine Virusinfektion von Birken auch in Mastjahren zu detektieren. Abbildung 8.10 zeigt den Vergleich von FTIR-Spektren CLRV-positiv und CLRV-negativ getesteter Pollenproben. Es treten die für Pollenproben charakteristischen Banden auf (siehe Tabelle 8.2), die u.a. Lipi-

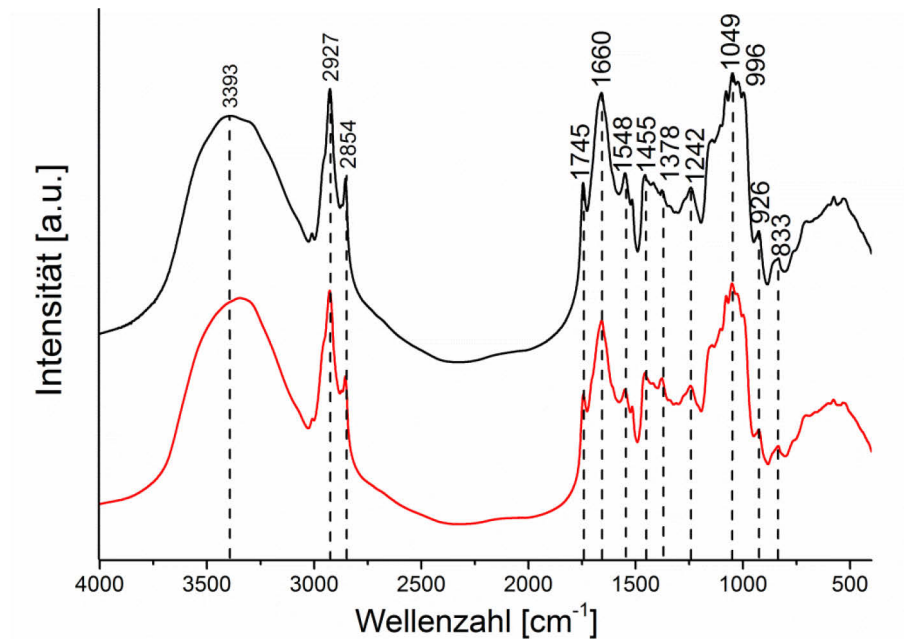


Abbildung 8.10: Vergleich von FTIR-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2012.

den, Proteinen und Kohlenhydraten zugeordnet werden können. Ähnlich wie bei der Untersuchung von Pollenproben des Jahres 2011 (vgl. Abbildung 8.4) können keine signifikanten Unterschiede von Spektren CLRV-positiv und CLRV-negativ getesteter Pollenproben identifiziert werden. Um die Varianzen des Datensatzes zu untersuchen, wurde anschließend eine PCA der Spektren durchgeführt (Abbildung 8.11). Anhand der Scores (Abbildung 8.11a) ist jedoch ebenfalls keine Identifizierung CLRV-positiver Pollenproben möglich. Betrachtet man die Loadings (Abbildung 8.11c), so ist erkennbar, dass diese sich signifikant von denen der PCA von Spektren aus dem Jahr 2011 unterscheiden (vgl. Abbildung 8.5b). Der größte Unterschied ist dabei, dass Spektren von Proben des Mastjahres 2012 bedeutend größere Unterschiede in der für Lipide charakteristischen Bande bei 1745 cm⁻¹ aufweisen. Dies spricht für hohe Variationen in der Lipidzusammensetzung, die nicht durch eine CLRV-Infektion verursacht werden. Zimmermann *et. al.* haben bereits gezeigt, dass Umweltparameter signifikant die Lipidzusammensetzung und damit auch lipidspezifische Banden im FTIR-Spektrum beeinflussen können.^[17] Es ist möglich, dass bei den hier untersuchten FTIR-Spektren Varianzen der chemischen Zusammensetzung von Pollenkörnern, die durch andere Umweltparameter hervorgerufen werden, die Varianzen überlagern, die durch die CLRV-Infektion verursacht werden. Dies könnte erklären, warum eine Identifizierung einer Virusinfektion mit FTIR-Spektroskopie für diese Proben scheitert.

Tabelle 8.2: Charakteristische Absorptionsbanden der FTIR-Spektren von Birkenpollenproben aus dem Jahr 2012 und deren Zuordnung zu Referenzspektren

Wellenzahl [cm^{-1}]	Zuordnung ^a
3393	ν , OH
2927	ν , CH_3 asym. ^c
2854	ν , CH_3 sym. ^c
1745	ν , C=O, Lipide ^b
1660	ν , C=O, Proteine, Amid I ^b
1548	ν , C-N, δ , N-H, Proteine, Amid II ^b
1455	δ , CH_2 , Lipide ^b
1378	δ , CH_2 , Kohlenhydrate
1242	ν , C-O, Lipide ^b
1049	ν , C-O-C, Kohlenhydrate ^{b,c}
996	ν , Kohlenhydrate ^{b,c}
926	Ringatmung, Kohlenhydrate ^{b,c}
833	Ringatmung, Sporopollenin ^b

^a Abkürzungen: ν , Streckschwingung δ , Deformationsschwingung
sym, symmetrisch asym, asymmetrisch

^b nach [16], [17]

^c nach [181]

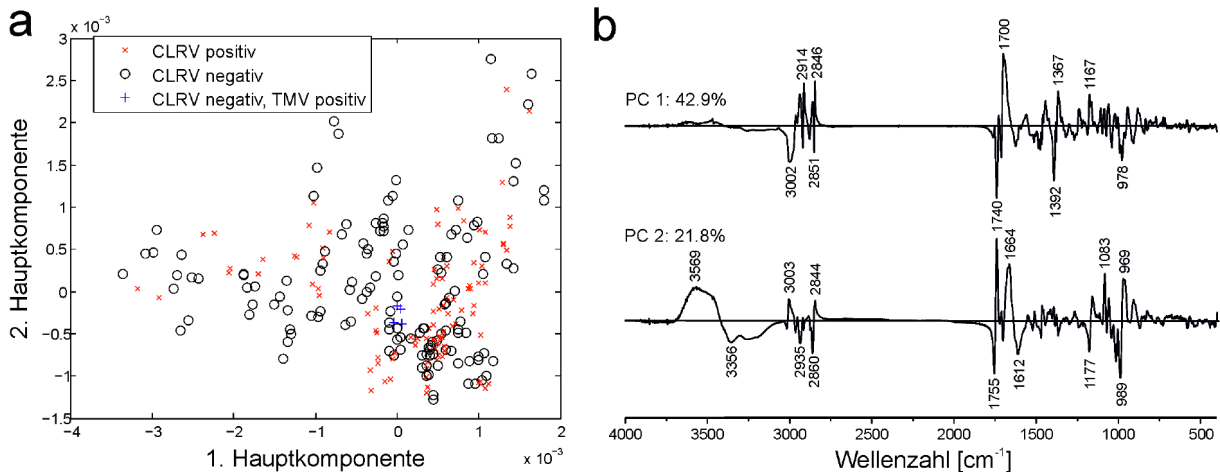


Abbildung 8.11: Ergebnis der Hauptkomponentenanalyse von 248 FTIR-Spektren zur Unterscheidung von Kaliumbromidpresslingen von CLRV-positiv und CLRV-negativ getesteten Birkenpollenproben aus dem Jahr 2012. Es sind die Scores der ersten und zweiten Hauptkomponente (a), sowie die Loadings und Varianzen der ersten beiden Hauptkomponenten (c) gezeigt.

8.3 Schlussfolgerungen

Die hier gezeigten Ergebnisse lassen vermuten, dass eine Virusinfektion von Birken durch die Untersuchung von Pollenproben mit MALDI-TOF MS von Säureextrakten und FTIR-Spektroskopie von Kaliumbromidpresslingen gelingt, wenn Proben untersucht werden, die in einem Jahr geringer Pollenproduktion gesammelt wurden. Da auch Pollenproben identifiziert werden konnten, bei denen eine Infektion mit dem Tabakmosaikvirus nachgewiesen wurde, könnte es möglich sein, durch diese Methoden eine unspezifische Virusinfektion der Pflanze und die damit einhergehende Änderung der chemischen Zusammensetzung nachzuweisen. Ob dies auch unspezifisch möglich ist, muss in zukünftigen Experimenten geklärt werden.

Aufgrund der Tatsache, dass es sich bei den verwendeten Methoden um komplementäre Methoden handelt, ist es sinnvoll, die Daten mit CPCA für eine robustere Klassifizierung zu kombinieren. Die Ergebnisse zeigen, dass durch die Virusinfektion Glykoprotein- und Lipidzusammensetzung verändert werden. Beides könnte sich auch auf das allergene Potential von Birkenpollen auswirken. Ein Zusammenhang dieser Art wurde bereits bei virusinfizierten Gräserpollen festgestellt.^[29]

Werden allerdings Pollenproben eines Mastjahres untersucht, ist es wahrscheinlich, dass die virusspezifischen Varianzen der chemischen Zusammensetzung der Pollenkörner mit solchen überlagern, die durch andere Umweltparameter verursacht werden. Eine Identifizierung der Virusinfektion ist bei diesen Pollenproben mit MALDI-TOF MS von Säureextrakten und FTIR-Spektroskopie von Kaliumbromidpresslingen nicht möglich. Die detaillierte Untersuchung der Wechselwirkung einer Virusinfektion mit anderen Umweltparametern auf die chemische Zusammensetzung von Pollenkörnern, sollte ebenfalls in zukünftigen Experimenten, auch unter Berücksichtigung zusätzlicher Kontrollexperimente, analysiert werden.

9. Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden neue analytische Methoden für die spektroskopische und spektrometrische Klassifizierung von Pollenproben angewendet. Dabei wurden folgende Verfahren erfolgreich mit multivariaten Auswertemethoden für eine taxonomische Klassifizierung kombiniert:

- *Matrix-assisted laser desorption/ionisation time-of-flight*-Massenspektrometrie (MALDI-TOF MS)
- Oberflächenverstärkte Raman-Streuung (*surface-enhanced Raman scattering*, SERS)
- Fourier-Transform-Infrarot-Spektroskopie (FTIR-Spektroskopie) von einzelnen Pollenkörnern

Die erhaltenen Daten wurden in einer *Consensus Principal Component Analysis* (CPCA) verknüpft.

Es wurde gezeigt, dass sowohl durch MALDI-TOF MS als auch durch SERS eine taxonomische Klassifizierung von Ordnungen, Gattungen, sowie von verschiedenen Arten innerhalb einer Gattung möglich ist. Damit ist eine genauere taxonomische Einteilung möglich als durch das aktuell verwendete Verfahren für die Pollenflugwarnung,^[11] wodurch das Potenzial dieser Methoden eindrücklich belegt wird.

Um die Identifizierung unbekannter Pollenproben zu ermöglichen, sollten in der Zukunft entsprechende Datenbanken aufgebaut werden, die MALDI-TOF MS- und SERS-Spektren von Pollenproben unterschiedlichster phylogenetischer Herkunft enthalten und mit Daten unbekannter Proben verglichen werden können. Auf Basis solcher Datenbanken könnten automatisierte und aktuelle Pollenflugwarnungen veröffentlicht werden. Auch Datenbanken mit Spektren unter Variation anderer, z.B. umweltspezifischer Parameter sind denkbar. Da sie auf Informationen der chemischen Zusammensetzung und nicht auf morphologischen Eigenschaften von Pollenkörnern beruhen, könnte es möglich sein, klimatische oder standortbedingte Einflüsse der Pflanze zu detektieren und z.B. Aussagen über Umweltbelastungen durch Abgase^[7] oder den Ernteertrag von Nutzpflanzen zu treffen.

Es wurde festgestellt, dass in MALDI-TOF MS-Spektren vor allem Unterschiede der Glykoproteinzusammensetzung der Pollenkörner sichtbar werden. Da es sich bei vielen Allergenen um Glykoproteine handelt, könnte eine weitere Anwendung dieser Methode auch in der Klassifizierung von Pollenproben mit unterschiedlicher Wirkung auf den menschlichen Organismus liegen. Dazu wird es allerdings notwendig sein, zusätzlich zur hier erreichten

Klassifizierung ganzer Spektren die Zuordnung von Banden zu spezifischen Molekülen weiter voranzutreiben.

Wie gezeigt wurde, beruht die Klassifizierung von SERS-Spektren wässriger Pollenextrakte auf spektralen Beiträgen der inneren, zellulären Bestandteile. Diese werden unabhängig von den äußeren Exinebestandteilen analysiert. Damit wurde eine neue, zu den bisherigen schwingungsspektroskopischen Verfahren^{[13],[165]} komplementäre Methode für die Untersuchung der komplexen, chemischen Zusammensetzung von Pollenkörnern erschlossen. Die erhaltenen SERS-Spektren beinhalten eine Vielzahl von Beiträgen unterschiedlicher Moleküle und Molekülklassen. Deshalb ist es eine zentrale Herausforderung bei der Untersuchung von Pollenproben durch diese Methode, die relevanten Informationen aus den komplexen Datenmengen zu extrahieren. Es wurde gezeigt, dass künstliche, neuronale Netze hierzu fähig sind und eine taxonomische Klassifizierung wässriger Pollenextrakte ermöglichen. Allgemein bedeutet dies für die Untersuchung komplexer Molekülgemische mit SERS, dass diese neben der Identifizierung einzelner Bestandteile^[81] auch als Ganze klassifiziert werden können. Die neuronalen Netze können auch dafür verwendet werden, die für die Klassifizierung relevanten Spektren aus dem komplexen Datensatz zu extrahieren und damit Datensätze zu erhalten, die die für eine taxonomische Klassifizierung relevanten Informationen enthalten.

In einer kombinierten Auswertung mit CPCA kann der MALDI-TOF MS-Block dazu verwendet werden, den SERS-Block zu überwachen und damit taxonomisch relevante Varianzen aus den SERS-Daten zu extrahieren. Da es sich bei MALDI-TOF MS und SERS um komplementäre Methoden handelt, ist die Kombination der Daten mit CPCA neben der Extraktion insbesondere auch für die Verbesserung der taxonomischen Klassifizierung interessant. Es kann dadurch ein sehr viel größerer Teil der chemischen Zusammensetzung komplexer, biologischer Proben in der Auswertung abgebildet und damit in die Untersuchung mit einbezogen werden. Die taxonomische Klassifizierung wird robuster und weniger anfällig für fehlerhafte Zuordnungen. Die Auswertung von Pollendaten durch CPCA würde eine Kombination weiterer komplementärer analytischer Methoden ermöglichen. So könnten beispielsweise zusätzlich zu Daten aus MALDI-TOF MS- und SERS-Untersuchungen auch solche, die durch FTIR-Spektroskopie von Kaliumbromidpresslingen^[16], Raman-Spektroskopie^[13] oder *laser-induced breakdown spectroscopy* (LIBS)^[24] entstehen, für eine kombinierte Auswertung verwendet werden. Das große Potenzial der Kombination verschiedener spektroskopischer Methoden durch CPCA für die Pollenanalytik wurde erst kürzlich anhand von Daten verschiedener FTIR- und Raman-Untersuchungen demonstriert.^[36]

Die Untersuchungen einzelner Pollenkörner mit FTIR-Transmissionsmessungen liefern Spektren, die eine Überlagerung von Absorptionsbanden mit Streuartefakten darstellen. Durch geeignete Spektrenvorbehandlung, die u.a. eine EMSC-Korrektur beinhaltet,^[163] wurde hier gezeigt, dass taxonomisch relevante Fingerabdrücke aus FTIR-Mikrospektren extrahiert werden können. Diese konnten in vielen Fällen korrekt klassifiziert werden. Im Fall einer zu großen Ähnlichkeit sowohl von Absorptionsspektren, als auch von Streuartefakten ist eine Klassifizierung nicht möglich. Um dies zu ändern, ist es notwendig, die Ursachen und den Einfluss der Streuartefakte noch detaillierter zu untersuchen und spezifische Algorithmen zu entwickeln, die aus den FTIR-Daten die Absorptionsmerkmale extrahieren können. Erste Ansätze in diesem Bereich sind bereits veröffentlicht.^[27] Diese könnten auch durch Kombination mit den in dieser Arbeit angewendeten Methoden die Klassifizierung einer größeren Anzahl an taxonomischen Gruppen ermöglichen.

Über die Entwicklung neuer Verfahren für die taxonomische Klassifizierung hinaus wurde ihre Anwendbarkeit zur Untersuchung *cherry leaf roll virus* (CLRV) infizierter Birkenpollen getestet. Während durch SERS des wässrigen Extraktes keine Unterschiede zwischen den Pollenproben gefunden wurden, konnten MALDI-TOF MS und FTIR-Spektroskopie zur Identifizierung einer Virusinfektion eingesetzt werden. Für eine automatische Identifizierung einer Virusinfektion könnten somit ebenfalls FTIR- und MALDI-TOF MS-Datenbanken angelegt werden, die verglichen mit aktuell angewendeten Methoden^[179] eine bedeutend schnellere Identifizierung einer Virusinfektion erlauben. Die Ergebnisse der Untersuchungen liefern Hinweise darauf, dass durch eine Virusinfektion Glykoprotein- und Lipidzusammensetzung von Pollenkörnern verändert werden. Dies lässt vermuten, dass sich eine Infektion auch auf das allergene Potential von Pollen auswirkt, was auch andere Untersuchungen nahe legen.^[29] Um die komplexen Auswirkungen einer Virusinfektion auf die chemische Zusammensetzung von Pollenkörnern jedoch möglichst vollständig zu untersuchen, sollten noch weitere komplementäre Methoden, wie beispielsweise die Raman-Spektroskopie einzelner Pollenkörner,^[13] oder die spektroskopische Untersuchung der Pollenkeimung^[18] für die Untersuchung eingesetzt und die Daten mit CPCA verknüpft werden. Dieses Beispiel illustriert sehr gut, dass es weiterhin notwendig sein wird, neue Methoden für die Untersuchung von Pollenproben zu erschließen und komplexe, biochemische Zusammenhänge durch das Verknüpfen der Daten komplementärer Methoden aufzudecken.

10. Anhang

10.1 Abkürzungsverzeichnis

Abkürzung	Erklärung
A	Adenin
ANN	<i>Artificial neural networks</i>
asym	Asymmetrisch
ATR	<i>Attenuated total reflection</i>
C	Cytosin
CCD	<i>Charge-coupled device</i>
cDNA	<i>Complementary Deoxyribonucleic acid</i>
CLRV	<i>Cherry leaf roll virus</i>
CPCA	<i>Consensus Principal Component Analysis</i>
Cys	Cystein
δ	Deformationsschwinung
DNA	<i>Deoxyribonucleic acid</i>
DTGS	Deuteriertes Triglycinsulfat
EMSC	<i>Extended Multiplicative Scatter Correction</i>
ESEM	<i>Environmental scanning electron microscopy</i>
FT	Fourier-Transform
G	Guanin
HCA	Hierarchische Clusteranalyse
His	Histidin
IC-RT-PCR	<i>Immunocapture reverse transcription-polymerase chain reaction</i>
IR	Infrarot
LIBS	<i>Laser-induced breakdown spectroscopy</i>
m/z	Masse-zu-Ladungsverhältnis
MALDI	<i>Matrix-assisted Laser Desorption/Ionization</i>
MCP	<i>Micro-channel plate</i>
MCT	<i>Mercury cadmium telluride</i>
Met	Methionin
MIR	Mittleres Infrarot
MS	Massenspektrometrie
ν	Streckschwingung
Nd:YAG-Laser	Neodym-dotierter Yttrium-Aluminium-Granat-Laser
NIR	Nahes Infrarot
PCA	<i>Principal Component Analysis</i>
Phe	Phenylalanin
PLSR	<i>Partial least squares regression</i>
rER	Raues Endoplasmatisches Retikulum
RNA	<i>Ribonucleic acid</i>
S	Sporopollenin
SEM	<i>Scanning electron microscopy</i>

Abkürzung	Erklärung
Ser	Serin
SERS	<i>Surface-enhanced Raman scattering</i>
SG	Savitzky-Golay
SVM	<i>Support vector machines</i>
sym	Symmetrisch
T	Thymin
TMV	Tabakmosaikvirus
TOF	<i>Time-of-flight</i>
Trp	Tryptophan
Tyr	Tyrosin
Val	Valin
WHC	<i>Water holding capacity</i>
WTA	<i>Winner-takes-all</i>

10.2 Abbildungsverzeichnis

Abbildung 1.1: Schematische Darstellung eines Pollenkorns nach [6].	1
Abbildung 2.1: Schematische Darstellung möglicher Prozesse bei der Wechselwirkung von einem Molekül mit Licht.	6
Abbildung 2.2: Clusteranalyse der Punkte 1-5 nach [67].	15
Abbildung 2.3: Matrizen der Hauptkomponentenanalyse nach [40].	16
Abbildung 2.4: Schematische Darstellung der mathematischen Grundlage der PLS1.	18
Abbildung 2.5: Schematische Darstellung des iterativen Verfahrens zur Bestimmung der CPCA-Parameter nach [74].	20
Abbildung 2.6: Beispiel für die Topologie eines vorwärtsgerichteten, zweischichtigen, neuronalen Netzes.	22
Abbildung 4.1: Repräsentative SERS-Spektren des wässrigen Pollenextraktes von <i>Secale cereale</i> -Pollen in Goldnanopartikellösung. Die Maßstabsskala entspricht 500 cps. Siehe Tabelle 4.1 für eine mögliche Zuordnung der charakteristischen Banden.	39
Abbildung 4.2: Mittelwertspektren von intakten, unbehandelten Birkenpollenkörnern (schwarz) und von Pollenkörnern des Raffinats (rot).	41
Abbildung 4.3: Ergebnis der PCA von Spektren von unbehandelten Birkenpollenkörnern und von Birkenpollenkörnern des Raffinats. Es sind die Scores der ersten und zweiten Hauptkomponente (a), der ersten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der entsprechenden Hauptkomponenten (c) gezeigt.	43
Abbildung 4.4: Ergebnis der PCA von SERS-Spektren wässriger Extrakte von kommerziell erhältlichen Pollenkörnern. Es sind die Scores der ersten und zweiten Hauptkomponente (a), die Scores der ersten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der entsprechenden Hauptkomponenten (c) gezeigt.	45

Abbildung 4.5: Ergebnis der PCA nach Transformation der Daten zu einer Standardnormalverteilung. Es sind die Scores (a), sowie die Loadings und Varianzen (b) der ersten beiden Hauptkomponenten gezeigt. 46

Abbildung 4.6: Mittelwertspektren von 500 vektornormierten SERS-Spektren des wässrigen Extraktes von: a; *Bassia scoparia*, b; *Artemisia absinthium*, c; *Artemisia tridentata*, d; *Iva xantifolia*, e; *Betula fontinalis*, f; *Juniperus scopulorum*, g; *Juglans nigra*, h; *Secale cereale*, i; *Sorghum halepense*, j; *Populus deltoides*, k; *Populus nigra*, l; *Populus tremuloides*, m; *Populus trichocarpa*, n; *Ulmus pumila*. 47

Abbildung 4.7: Ergebnis der hierarchischen Clusteranalyse von 112 Mittelwertspektren (jeweils acht Mittelwertspektren von 14 verschiedenen Pollenarten) im Spektralbereich zwischen 400 und 1700 cm^{-1} . Es wurden jeweils 500 vektornormierte, abgeleitete Spektren gemittelt und die spektralen Daten zu einer Standardnormalverteilung transformiert. Zur besseren Übersichtlichkeit sind die Spektren von jeweils einer Art in gleichen Farben, aber anderer Nummerierung dargestellt. Die Nummerierung kennzeichnet verschiedene Cluster derselben Art, die häufig den einzelnen Extrakten, die jeweils für jede Art untersucht wurden, entsprechen. 50

Abbildung 4.8: Ergebnis der PCA von 112 Mittelwertspektren (jeweils acht Mittelwertspektren von 14 verschiedenen Pollenarten) im Spektralbereich zwischen 400 und 1700 cm^{-1} . Loadings der ersten vier Hauptkomponenten (a), sowie die Scores der ersten und zweiten (b), ersten und dritten (c) und ersten und vierten Hauptkomponente (d) sind gezeigt. Es wurden jeweils 500 vektornormierte, abgeleitete Spektren gemittelt und die spektralen Daten zu einer Standardnormalverteilung transformiert. 51

Abbildung 4.9: Ergebnis der ANN-Klassifizierung von SERS-Daten von 14 verschiedenen, kommerziell erhältlichen Pollenarten. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 53

Abbildung 4.10: Ergebnis der ANN-Klassifizierung von 110 Datensätzen der Ordnung *Coniferales* und 82 Datensätzen der Ordnung *Fagales*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 54

Abbildung 4.11: Ergebnis der ANN-Klassifizierung von 16 Datensätzen der Gattung *Taxus*, 26 Datensätzen der Gattung *Picea*, jeweils sechs Datensätzen der Gattungen *Juniperus* und *Abies*, sowie 56 Datensätzen der Gattung *Pinus*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 54

Abbildung 4.12: Ergebnis der ANN-Klassifizierung von 18 Datensätzen der Gattung *Alnus*, 20 Datensätzen der Gattung *Betula*, 20 Datensätzen der Gattung *Quercus*, sechs Datensätze der Gattung *Myrica*, sowie 18 Datensätzen der Gattung *Corylus*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 55

Abbildung 4.13: Ergebnis der ANN-Klassifizierung von 14 Datensätzen der Art *Pinus mugo* und zwölf Datensätzen der Art *Pinus sylvestris*. Es wurden jeweils die Zuordnungen von 500 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 56

Abbildung 4.14: Schematische Darstellung der angewendeten Extraktionen. 57

Abbildung 4.15: Ergebnis der Clusteranalyse von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. 58

Abbildung 4.16: Ergebnis der PCA von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Die Scores der ersten beiden Hauptkomponenten (a), sowie die Loadings der ersten Hauptkomponente (b) sind gezeigt. 58

Abbildung 4.17: Ergebnis der Clusteranalyse von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. 59

Abbildung 4.18: Ergebnis der PCA von 360 Mittelwertspektren, die aus für die Ordnung als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Die Scores der ersten beiden Hauptkomponenten (a), sowie die Loadings der ersten Hauptkomponente (b) sind gezeigt. 60

Abbildung 4.19: Ergebnis der Clusteranalyse von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a) und *Fagales* (b) als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. 62

Abbildung 4.20: Ergebnisse der PCA von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a-c) und *Fagales* (d-f) als taxonomisch spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Es sind die Scores der ersten, zweiten, dritten und vierten Hauptkomponente für Proben der Ordnung *Coniferales* (a+b), die Scores der ersten, zweiten, dritten und fünften Hauptkomponente für Proben der Ordnung *Fagales* (d+e), sowie die entsprechenden Loadings der jeweiligen Hauptkomponenten (c+e) gezeigt. 63

Abbildung 4.21: Ergebnis der Clusteranalyse von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a) und *Fagales* (b) als spezifisch identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. 64

Abbildung 4.22: Ergebnisse der PCA von Mittelwertspektren, die aus für die Gattungen der Ordnung *Coniferales* (a-c) und *Fagales* (d-f) als taxonomisch relevant identifizierten Spektren gebildet und anschließend zu einer Standardnormalverteilung transformiert wurden. Es sind die Scores der ersten, zweiten, dritten und vierten Hauptkomponente für Proben der Ordnung *Coniferales* (a+b), die Scores der ersten, zweiten, dritten und fünften Hauptkomponente für Proben der Ordnung *Fagales* (d+e), sowie die entsprechenden Loadings der jeweiligen Hauptkomponenten (c+e) gezeigt. 67

Abbildung 5.1: MALDI-TOF MS-Spektren kommerziell erhältlicher Pollenproben im Bereich zwischen m/z 1000 und 20000, bzw. zwischen m/z 4000 und 11000 (vergrößerter Bereich in grau). 71

Abbildung 5.2: Ergebnis der Klassifizierung von 240 Spektren der 10 kommerziell erhältlichen Pollenproben mit HCA. Es wurden euklidische Distanzen im Massebereich zwischen m/z 1000 und 15000 und der Ward's Algorithmus verwendet. 72

Abbildung 5.3: Charakteristische MALDI-TOF MS-Spektren ausgewählter Pollenproben der Ordnungen *Coniferales* (a) und *Fagales* (b). 73

Abbildung 5.4: Ergebnis der Klassifizierung von 188 Pollenspektren von 74 verschiedenen Arten der Ordnungen *Coniferales* und *Fagales* mit HCA. Es wurden euklidische Distanzen im Massebereich zwischen 1000 und 10000 m/z und der Ward's Algorithmus verwendet. 74

Abbildung 5.5: MALDI-TOF MS-Spektren von *Alnus firma* (a), *Alnus cordata* (b), *Alnus subcordata* (c), und *Abies concolor* (d+e). 75

Abbildung 5.6: Ergebnis der HCA von 97 MALDI-TOF MS-Pollenspektren der Ordnung *Coniferales* (a) und von 91 MALDI-TOF MS-Spektren der Ordnung *Fagales* (b). Es wurden euklidische Distanzen im Massebereich zwischen 1000 und 10000 m/z und der Ward's Algorithmus verwendet. 76

Abbildung 5.7: Ergebnis der PCA von 188 MALDI-TOF MS-Pollenspektren der Ordnungen *Coniferales* und *Fagales*. Es sind die Scores der ersten beiden Hauptkomponenten (a) und die Loadings und Varianz der ersten Hauptkomponente (b) gezeigt. 77

Abbildung 5.8: Ergebnisse der Hauptkomponentenanalysen von 97 MALDI-TOF MS-Pollenspektren der Ordnung *Coniferales* (a-c) und von 91 MALDI-TOF MS-Pollenspektren der Ordnung *Fagales* (d-f). Es sind die Scores der ersten, zweiten und vierten Hauptkomponente für Proben der Ordnung *Coniferales* (a+b), die Scores der ersten, zweiten und dritten Hauptkomponente für Proben der Ordnung *Fagales* (d+e), sowie die entsprechenden Loadings und Varianzen der jeweiligen Hauptkomponenten (c+e) gezeigt. 79

Abbildung 5.9: Ergebnisse der PCA von Pollenspektren verschiedener Gattungen der Ordnung *Coniferales*. Scores der ersten und zweiten Hauptkomponente für *Picea* (a), *Abies* (b), *Juniperus* (c), *Pinus* (d), und *Taxus* (e), sowie die entsprechenden Loadings und Varianzen der jeweiligen Hauptkomponenten (f-j) sind gezeigt. 82

Abbildung 5.10: Ergebnisse der PCA von Pollenspektren verschiedener Gattungen der Ordnung *Fagales*. Scores der ersten und zweiten Hauptkomponente für *Alnus* (a), *Corylus* (c), *Myrica* (d) und *Quercus* (e), Scores für die erste und vierte Hauptkomponente für *Betula* (b), sowie die entsprechenden Loadings und Varianzen der jeweiligen Hauptkomponenten (f-j) sind gezeigt. 83

Abbildung 6.1: Ergebnis der CPCA von 180 SERS- und MALDI-TOF MS-Datensätzen zur Klassifizierung auf taxonomischer Ebene der Ordnung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), sowie die Loadings der ersten beiden CPCA-Komponenten (d) sind gezeigt. 86

Abbildung 6.2: Ergebnis der CPCA von 180 Datensätzen nicht selektierter SERS- und MALDI-TOF MS-Spektren zur Klassifizierung auf taxonomischer Ebene der Ordnung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), sowie die Loadings der ersten beiden CPCA-Komponenten (d) sind gezeigt. 88

Abbildung 6.3: Ergebnisse der CPCA von 95 SERS- und MALDI-TOF MS-Datensätzen der Ordnung *Coniferales* zur Klassifizierung auf taxonomischer Ebene der Gattung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponenten für die einzelnen Blöcke (a+b) und den globalen Block (c), die Scores und Varianzen der dritten und vierten CPCA-Komponenten für die einzelnen Blöcke (d+e) und den globalen Block (f), sowie die Loadings der jeweiligen CPCA-Komponenten (g) sind gezeigt. 90

Abbildung 6.4: Ergebnisse der CPCA von 85 SERS- und MALDI-TOF MS-Datensätzen der Ordnung *Fagales* zur Klassifizierung auf Ebene der taxonomischen Gattung. Die Scores und Varianzen der ersten und zweiten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), die Scores und Varianzen der dritten und vierten CPCA-Komponente für die einzelnen Blöcke (d+e) und den globalen Block (f), sowie die Loadings der jeweiligen CPCA-Komponenten (g) sind gezeigt. 92

Abbildung 7.1: Repräsentative FTIR-Spektren einzelner Pollenkörner der Arten *Juniperus scopulorum* (a), *Populus nigra* (b) und *Secale cereale* (c). Für eine Zuordnung der Absorptionsbanden siehe Tabelle 7.1. 97

Abbildung 7.2: Varianzbasiertes, hierarchisches Klassifizierungsmodell für die Klassifizierung von FTIR-Spektren einzelner Pollenkörner von kommerziell erhältlichen Pollenproben. Die Optimierung der Anzahl an SG-Datenpunkten für die nummerierten Knotenpunkte kann der folgenden Abbildung 7.3 entnommen werden. 99

Abbildung 7.3: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte 1 bis 13 des hierarchischen Klassifizierungsmodells in Abbildung 7.2. Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt. 101

Abbildung 7.4: Regressionskoeffizienten der Knotenpunkte 1 bis 13 des hierarchischen Klassifizierungsmodells (siehe Abbildung 7.2), welches durch Training mit 140 Pollenspektren erhalten wurde. 102

Abbildung 7.5: Ergebnisse der PLSR-Klassifizierung von 199 Spektren kommerziell erhältlicher Pollenproben nach hierarchischem Klassifizierungsmodell (siehe Abbildung 7.2), welches durch Training mit 140 Pollen-spektren erhalten wurde. 103

Abbildung 7.6: Optimierte, hierarchisches Klassifizierungsmodell für die Klassifizierung von FTIR-Spektren einzelner Pollenkörner von kommerziell erhältlichen Pollenproben. Es wurden zusätzlich zum hierarchischen Modell in Abbildung 7.2 sieben Knotenpunkte eingefügt (I-VII), die der Kontrolle systematischer Fehlklassifizierung dienen. Die Optimierung der Anzahl an SG-Datenpunkten für die zusätzlichen Knotenpunkte können der folgenden Abbildung 7.7 entnommen werden. 104

Abbildung 7.7: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte I bis VII des hierarchischen Klassifizierungsmodells in Abbildung 7.6. Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt. 105

Abbildung 7.8: Regressionskoeffizienten der Knotenpunkte I bis VII des optimierten, hierarchischen Klassifizierungsmodells (siehe Abbildung 7.6), welches durch Training mit 140 Pollenspektren erhalten wurde. 106

Abbildung 7.9: Ergebnisse der PLSR-Klassifizierung von 199 Spektren kommerziell erhältlicher Pollenproben nach optimiertem hierarchischem Klassifizierungsmodell (siehe Abbildung 7.6), welches durch Training mit 140 Pollenspektren erhalten wurde. 107

Abbildung 7.10: Ausgewählte FTIR-Spektren einzelner Pollenkörner der Ordnungen *Coniferales* (links) und *Fagales* (rechts). Es sind Spektralbereiche gekennzeichnet, in denen charakteristische Banden von Lipiden und Proteinen detektiert werden können. Sporopollenin-spezifische Banden sind zusätzlich mit einem S gekennzeichnet. 108

Abbildung 7.11: ESEM Aufnahmen von Pollenkörnern der Ordnung *Coniferales* und deren geschätzte Größe. Die *Sacci* der Pollenkörner der Familie *Pinaceae* sind mit einem S gekennzeichnet. 109

Abbildung 7.12: ESEM Aufnahmen von Pollenkörnern der Ordnung *Fagales* und deren geschätzte Größe. 110

Abbildung 7.13: Hierarchisches Klassifizierungsmodell für die taxonomische Klassifizierung von FTIR-Spektren einzelner Pollenkörner von Pollenproben aus dem Feld. Die Optimierung der Anzahl an SG-Datenpunkten für die nummerierten Knotenpunkte kann der folgenden Abbildung 7.14 entnommen werden. 111

Abbildung 7.14: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte 1 bis 8 des hierarchischen Klassifizierungsmodells (siehe Abbildung 7.13). Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt. 113

Abbildung 7.15: Regressionskoeffizienten der Knotenpunkte 1 bis 8 des hierarchischen Klassifizierungsmodells (siehe Abbildung 7.13), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde. Sporopollenin-spezifische Banden sind mit einem S gekennzeichnet. 114

Abbildung 7.16: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Ordnung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach hierarchischem Klassifizierungsmodell (siehe Abbildung 7.13), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert. 115

Abbildung 7.17: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Gattung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach hierarchischem Klassifizierungsmodell (siehe Abbildung 7.13), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert. 115

Abbildung 7.18: Optimierte, hierarchisches Klassifizierungsmodell für die Klassifizierung von FTIR-Spektren einzelner Pollenkörner von Pollenproben aus dem Feld. Es wurden zusätzlich zum hierarchischen Modell in Abbildung 7.13 sechs Knotenpunkte eingefügt (I-VI), die der Kontrolle systematischer Fehlklassifizierung dienen. Die Optimierung der Anzahl an SG-Datenpunkten für die zusätzlichen Knotenpunkte können der folgenden Abbildung 7.19 entnommen werden. 116

Abbildung 7.19: Optimierung der Anzahl an SG-Datenpunkten bei der zweiten Ableitung gefolgt von einer EMSC-Korrektur der Pollenkornspektren für die Knotenpunkte I bis VI des hierarchischen Klassifizierungsmodells in Abbildung 7.18. Es sind jeweils die Anzahl an korrekt klassifizierten Spektren (a), das Bestimmtheitsmaß R^2 (b) und die Anzahl verwendeter PLSR-Komponenten (c) in Abhängigkeit der Anzahl an SG-Datenpunkten gezeigt. 117

Abbildung 7.20: Regressionskoeffizienten der Knotenpunkte I bis VI des optimierten hierarchischen Klassifizierungsmodells (siehe Abbildung 7.18), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde. Sporopollenin-spezifische Banden sind mit einem S gekennzeichnet. 118

Abbildung 7.21: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Ordnung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach optimiertem hierarchischem Klassifizierungsmodell (siehe Abbildung 7.18), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert. 119

Abbildung 7.22: Ergebnisse der PLSR-Klassifizierung auf taxonomischer Ebene der Gattung. Es wurden 784 Spektren von 39 verschiedenen Pollenproben aus dem Feld nach optimiertem hierarchischem Klassifizierungsmodell (siehe Abbildung 7.18), welches durch Training mit 768 Spektren von 38 Pollenproben erhalten wurde, klassifiziert. 120

Abbildung 8.1: Ergebnis der ANN-Klassifizierung von 46 SERS-Datensätzen von Extrakten aus Birkenpollen aus dem Jahr 2011. Es wurden jeweils die Zuordnungen von 1000 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 125

Abbildung 8.2: Vergleich repräsentativer MALDI-TOF MS-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2011. Es ist jeweils der Massenbereich zwischen m/z 4400 und 8000, sowie die vergrößerten Massenbereiche zwischen m/z 6760 und 6860 und m/z 7080 und 7180 (in grau) gezeigt. 126

Abbildung 8.3: Ergebnis der PCA von 82 MALDI-TOF MS-Spektren zur Unterscheidung CLRV-positiv und CLRV-negativ getesteter Birkenpollenproben aus dem Jahr 2011. Es sind die Scores der ersten und zweiten Hauptkomponente (a), die der zweiten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der ersten drei Hauptkomponenten (c) gezeigt. 127

Abbildung 8.4: Vergleich repräsentativer FTIR-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2011. 128

Abbildung 8.5: Ergebnis der PCA von 82 FTIR-Spektren zur Unterscheidung von Kaliumbromidpresslingen von CLRV-positiv und CLRV-negativ getesteten Birkenpollenproben aus dem Jahr 2011. Es sind die Scores der ersten und zweiten Hauptkomponente (a), sowie die Loadings und Varianzen der ersten beiden Hauptkomponenten (c) gezeigt. 130

Abbildung 8.6: Ergebnis der CPCA von MALDI-TOF MS- und FTIR-Daten zur Unterscheidung von 82 Datensätzen in CLRV-positiv und CLRV-negativ getestete Birkenpollenproben aus dem Jahr 2011. Die Scores und Varianzen der ersten und dritten CPCA-Komponente für die einzelnen Blöcke (a+b) und den globalen Block (c), sowie die Loadings der ersten und dritten CPCA-Komponente (d) sind gezeigt. 131

Abbildung 8.7: Ergebnis der ANN-Klassifizierung von 84 Datensätzen von Extrakten aus Birkenpollen aus dem Jahr 2012. Es wurden jeweils die Zuordnungen von 1000 Spektren gemittelt und eine Klassifizierung mit winner-takes-all (WTA) Methode durchgeführt. 133

Abbildung 8.8: Vergleich von MALDI-TOF MS-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2012. Es ist jeweils der Massenbereich zwischen m/z 4400 und 8000, sowie die vergrößerten Massenbereiche zwischen m/z 6760 und 6860 und m/z 7080 und 7180 (in grau) gezeigt. 134

Abbildung 8.9: Ergebnis der PCA von 124 MALDI-TOF MS-Spektren zur Unterscheidung CLRV-positiv und CLRV-negativ getesteter Birkenpollenproben aus dem Jahr 2012. Es sind die Scores der ersten und zweiten Hauptkomponente (a), die der zweiten und dritten Hauptkomponente (b), sowie die Loadings und Varianzen der ersten drei Hauptkomponenten (c) gezeigt. 135

Abbildung 8.10: Vergleich von FTIR-Spektren CLRV-negativ (schwarz) und CLRV-positiv (rot) getesteter Birkenpollenproben aus dem Jahr 2012. 136

Abbildung 8.11: Ergebnis der Hauptkomponentenanalyse von 248 FTIR-Spektren zur Unterscheidung von Kaliumbromidpresslingen von CLRV-positiv und CLRV-negativ getesteten Birkenpollenproben aus dem Jahr 2012. Es sind die Scores der ersten und zweiten Hauptkomponente (a), sowie die Loadings und Varianzen der ersten beiden Hauptkomponenten (c) gezeigt. 137

10.3 Tabellenverzeichnis

Tabelle 3.1: Die in dieser Arbeit untersuchten, kommerziell erhältlichen Pollenproben	23
Tabelle 3.2: Die in dieser Arbeit untersuchten Pollenproben aus dem Feld	24
Tabelle 3.3: Die in dieser Arbeit untersuchten Birkenpollenproben und Ergebnisse der IC-RT-PCR	26
Tabelle 4.1: Raman-Verschiebungen charakteristischer Banden in den SERS-Spektren und deren mögliche Zuordnung zu Referenzspektren	40
Tabelle 4.2: Raman-Verschiebungen charakteristischer Banden in den Raman-Spektren und deren mögliche Zuordnung zu Referenzspektren	42
Tabelle 4.3: Raman-Verschiebungen charakteristischer Banden in den SERS-Mittelwertspektren und deren mögliche Zuordnung zu Referenzspektren	48
Tabelle 7.1: Charakteristische Absorptionsbanden der FTIR-Spektren einzelner Pollenkörner und deren Zuordnung zu Referenzspektren	98
Tabelle 8.1: Charakteristische Absorptionsbanden der FTIR-Spektren von Birkenpollenproben aus dem Jahr 2011 und deren Zuordnung zu Referenzspektren	129
Tabelle 8.2: Charakteristische Absorptionsbanden der FTIR-Spektren von Birkenpollenproben aus dem Jahr 2012 und deren Zuordnung zu Referenzspektren	137

10.4 Literaturverzeichnis

- [1] Gould K. S., L. C., *Flavonoids: Chemistry, Biochemistry and Applications* CRC Press Taylor & Francis Group Boca Raton, Florida, USA, 2006.
- [2] Brooks, J.; Shaw, G. *Nature*, **1968**, *219*, 532-533.
- [3] Scott, R. W.; Strohl, M. J. *Phytochemistry*, **1962**, *1*, 189-193.
- [4] Ariizumi, T.; Toriyama, K. *Annu. Rev. Plant Biol.*, **2011**, *62*, 437-460.
- [5] Pacini, E.; Hesse, M. *Flora*, **2005**, *200*, 399-415.
- [6] Piffanelli, P.; Ross, J. H. E.; Murphy, D. J. *Sex. Plant Reprod.*, **1998**, *11*, 65-80.
- [7] Behrendt, H.; Becker, W.-M. *Curr. Opin. Immunol.*, **2001**, *13*, 709-715.
- [8] Ring, J.; Krämer, U.; Schäfer, T.; Behrendt, H. *Curr. Opin. Immunol.*, **2001**, *13*, 701-708.
- [9] Jackson, M. *Clin. Exp. Allergy*, **2001**, *31*, 1665-1671.
- [10] Barnett, S. B. L.; Nurmagambetov, T. A. *J. Allergy Clin. Immunol.*, **2011**, *127*, 145-152.
- [11] Deutscher Wetterdienst. <http://www.dwd.de/>.
- [12] Pappas, C. S.; Tarantilis, P. A.; Harizanis, P. C.; Polissiou, M. G. *Appl. Spectrosc.*, **2003**, *57*, 23-27.
- [13] Schulte, F.; Lingott, J.; Panne, U.; Kneipp, J. *Anal. Chem.*, **2008**, *80*, 9551-9556.
- [14] Pummer, B. G.; Bauer, H.; Bernardi, J.; Chazallon, B.; Facq, S.; Lendl, B.; Whitmore, K.; Grothe, H. *J. Raman Spectrosc.*, **2013**, *44*, 1654-1658.
- [15] Schulte, F.; Mader, J.; Kroh, L. W.; Panne, U.; Kneipp, J. *Anal. Chem.*, **2009**, *81*, 8426-8433.
- [16] Zimmermann, B. *Appl. Spectrosc.*, **2010**, *64*, 1364-1373.
- [17] Zimmermann, B.; Kohler, A. *PLoS ONE*, **2014**, *9*, e95417.
- [18] Schulte, F.; Panne, U.; Kneipp, J. *J. Biophotonics*, **2010**, *3*, 542-547.
- [19] Joseph, V.; Gensler, M.; Seifert, S.; Gernert, U.; Rabe, J. P.; Kneipp, J. *J. Phys. Chem. C*, **2012**, *116*, 6859-6865.
- [20] Joseph, V.; Schulte, F.; Roach, H.; Feldmann, I.; Dorfel, I.; Osterle, W.; Panne, U.; Kneipp, J. *Chem. Commun.*, **2011**, *47*, 3236-3238.

- [21] Seifert, S. *Diplomarbeit, Humboldt-Universität zu Berlin*, **2011**.
- [22] Laucks, M. L.; Roll, G.; Schweiger, G.; Davis, E. J. *J. Aerosol Sci*, **2000**, *31*, 307-319.
- [23] Ivleva, N.; Niessner, R.; Panne, U. *Anal. Bioanal. Chem.*, **2005**, *381*, 261-267.
- [24] Boyain-Goitia, A. R.; Beddows, D. C. S.; Griffiths, B. C.; Telle, H. H. *Appl. Opt.*, **2003**, *42*, 6119-6132.
- [25] Kneipp, J.; Kneipp, H.; Kneipp, K. *Chem. Soc. Rev.*, **2008**, *37*, 1052-1060.
- [26] Zimmermann, B.; Tkalc̃ec, Z.; Mešić, A.; Kohler, A. *PLoS ONE*, **2015**, *10*, e0124240.
- [27] Lukacs, R.; Blumel, R.; Zimmerman, B.; Bagcioglu, M.; Kohler, A. *Analyst*, **2015**, *140*, 3273-3284.
- [28] Lasch, P.; Beyer, W.; Nattermann, H.; Stammeler, M.; Siegbrecht, E.; Grunow, R.; Naumann, D. *Appl. Environ. Microbiol.*, **2009**, *75*, 7229-7242.
- [29] Pallett, D. W.; Soh, E.; Edwards, M. L.; Bodey, K.; Lau, L. C.; Cooper, J. I.; Howarth, P. H.; Walls, A. F.; Wang, H. *Environ. Health*, **2009**, *8*.
- [30] Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C. W.; Tsai, J. C.; Kang, J. X.; Xie, X. S. *Science*, **2008**, *322*, 1857-1861.
- [31] Bernhard, S. *Angew. Chem.*, **1973**, *85*, 925-950.
- [32] Günzler, H.; Heise, H. M., *IR-Spektroskopie* VCH Verlagsgesellschaft: 1996.
- [33] Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A.; Hughes, C.; Lasch, P.; Martin-Hirsch, P. L.; Obinaju, B.; Sockalingum, G. D.; Sule-Suso, J.; Strong, R. J.; Walsh, M. J.; Wood, B. R.; Gardner, P.; Martin, F. L. *Nat. Protoc.*, **2014**, *9*, 1771-1791.
- [34] Mantsch, H. H.; Choo-Smith, L. P.; Shaw, R. A. *Vib. Spectrosc*, **2002**, *30*, 31-41.
- [35] Udelhoven, T.; Naumann, D.; Schmitt, J. *Appl. Spectrosc.*, **2000**, *54*, 1471-1479.
- [36] Bagcioglu, M.; Zimmermann, B.; Kohler, A. *PLoS ONE*, **2015**, *10*.
- [37] Carr, G. L.; Reffner, J. A.; Williams, G. P. *Rev. Sci. Instrum.*, **1995**, *66*, 1490-1492.
- [38] Wetzel, D. L.; LeVine, S. M. *Science*, **1999**, *285*, 1224-1225.
- [39] Martens, H.; Stark, E. *J. Pharm. Biomed. Anal.*, **1991**, *9*, 625-635.
- [40] Kessler, W., *Multivariate Datenanalyse*. 1. ed.; Wiley-VCH: **2007**.
- [41] Bassan, P.; Byrne, H. J.; Bonnier, F.; Lee, J.; Dumas, P.; Gardner, P. *Analyst*, **2009**, *134*, 1586-1593.

- [42] Kohler, A.; Sule-Suso, J.; Sockalingum, G. D.; Tobin, M.; Bahrami, F.; Yang, Y.; Pijanka, J.; Dumas, P.; Cotte, M.; van Pittius, D. G.; Parkes, G.; Martens, H. *Appl. Spectrosc.*, **2008**, 62, 259-266.
- [43] Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Byrne, H. J.; Dumas, P.; Gazi, E.; Brown, M.; Clarke, N.; Gardner, P. *Analyst*, **2009**, 135, 268-277.
- [44] Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Byrne, H. J.; Dumas, P.; Gazi, E.; Brown, M.; Clarke, N.; Gardner, P. *Analyst*, **2010**, 135, 268-277.
- [45] Bassan, P.; Kohler, A.; Martens, H.; Lee, J.; Jackson, E.; Lockyer, N.; Dumas, P.; Brown, M.; Clarke, N.; Gardner, P. *J. Biophotonics*, **2010**, 3, 609-620.
- [46] Smekal, A. *Naturwissenschaften*, **1923**, 11, 873-875.
- [47] Raman, C. V.; Krishnan, K. S. *Nature*, **1928**, 121, 501-502.
- [48] Landsberg, G.; Mandelstam, L. *Zeitschrift Fur Physik*, **1928**, 50, 769-780.
- [49] Baena, J. R.; Lendl, B. *Curr. Opin. Chem. Biol.*, **2004**, 8, 534-539.
- [50] McCreery, R. L., *Raman Spectroscopy for Chemical Analysis*. 1 ed.; Wiley-Interscience: **2000**.
- [51] Fleischmann, M.; Hendra, P. J.; Mcquilla. *Chem. Phys. Lett.*, **1974**, 26, 163-166.
- [52] Albrecht, M. G.; Creighton, J. A. *J. Am. Chem. Soc.*, **1977**, 99, 5215-5217.
- [53] Jeanmaire, D. L.; Van Duyne, R. P. *J. Electroanal. Chem.*, **1977**, 84, 1-20.
- [54] Zhao, L. L.; Jensen, L.; Schatz, G. C. *J. Am. Chem. Soc.*, **2006**, 128, 2911-2919.
- [55] Otto, A.; Billmann, J.; Eickmans, J.; Erturk, U.; Pettenkofer, C. *Surf. Sci.*, **1984**, 138, 319-338.
- [56] Pettinger, B. *Chem. Phys. Lett.*, **1984**, 110, 576-581.
- [57] Kneipp, K.; Kneipp, H.; Itzkan, I.; Dasari, R. R.; Feld, M. S. *J. Phys.-Condens. Mat.*, **2002**, 14, R597-R624.
- [58] Shanmukh, S.; Jones, L.; Zhao, Y. P.; Driskell, J. D.; Tripp, R. A.; Dluhy, R. A. *Anal. Bioanal. Chem.*, **2008**, 390, 1551-1555.
- [59] Jarvis, R. M.; Goodacre, R. *Anal. Chem.*, **2004**, 76, 40-47.
- [60] Gross, J. H., *Massenspektrometrie*. Springer Spektrum Heidelberg, **2013**.
- [61] McDonnell, L. A.; Heeren, R. M. A. *Mass Spectrom. Rev.*, **2007**, 26, 606-643.

- [62] Hettick, J. M.; Green, B. J.; Buskirk, A. D.; Kashon, M. L.; Slaven, J. E.; Janotka, E.; Blachere, F. M.; Schmechel, D.; Beezhold, D. H. *Anal. Biochem.*, **2008**, *380*, 276-281.
- [63] Holland, R. D.; Wilkes, J. G.; Rafii, F.; Sutherland, J. B.; Persons, C. C.; Voorhees, K. J.; Lay, J. O. *Rapid Commun. Mass Spectrom.*, **1996**, *10*, 1227-1232.
- [64] Yssouf, A.; Flaudrops, C.; Drali, R.; Kernif, T.; Socolovski, C.; Berenger, J. M.; Raoult, D.; Parola, P. *J. Clin. Microbiol.*, **2013**, *51*, 522-528.
- [65] Danilatos, G. D. *Advances in Electronics and Electron Physics*, **1988**, *71*, 109-250.
- [66] Collins, S. P.; Pope, R. K.; Scheetz, R. W.; Ray, R. I.; Wagner, P. A.; Little, B. J. *Microsc. Res. Tech.*, **1993**, *25*, 398-405.
- [67] *The MathWorks Inc., Matlab*, **2008**, 7.6.0.324 (R2008a).
- [68] Pearson, K. *Philosophical Magazine*, **1901**, *2*, 559-572.
- [69] Wold, H. *European Economic Review*, **1974**, *5*, 67-86.
- [70] Gerlach, R. W.; Kowalski, B. R.; Wold, H. O. A. *Anal. Chim. Acta*, **1979**, *112*, 417-421.
- [71] Martens, H.; Naes, T., *Multivariate Calibration*. Wiley: Chichester, **1989**.
- [72] Perisic, N.; Afseth, N. K.; Ofstad, R.; Hassani, S.; Kohler, A. *Food Chem.*, **2013**, *138*, 679-686.
- [73] Perisic, N.; Afseth, N. K.; Ofstad, R.; Narum, B.; Kohler, A. *Meat Sci.*, **2013**, *95*, 576-585.
- [74] Hassani, S. *Dissertation, Norwegian University of life sciences*, **2012**.
- [75] McCulloch, W. S.; Pitts, W. *Bull. Math. Biophys.*, **1943**, *5*, 115-133.
- [76] Pitts, W.; McCulloch, W. S. *Bull. Math. Biophys.*, **1947**, *9*, 127-147.
- [77] Rosenblatt, F. *Psychol. Rev.*, **1958**, *65*, 386-408.
- [78] Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature*, **1986**, *323*, 533-536.
- [79] Werbos, P. J. *Proceedings of the Ieee*, **1990**, *78*, 1550-1560.
- [80] Zell, A., *Simulation neuronaler Netze* Addison Wesley Longman Verlag GmbH: München **1997**.
- [81] Kasera, S.; Herrmann, L. O.; Barrio, J. d.; Baumberg, J. J.; Scherman, O. A. *Sci. Rep.*, **2014**, *4*.
- [82] Lee, P. C.; Meisel, D. *J. Phys. Chem.*, **1982**, *86*, 3391-3395.

- [83] Nabiev, I. R.; Morjani, H.; Manfait, M. *Eur. Biophys. J.*, **1991**, *19*, 311-316.
- [84] Kneipp, K.; Haka, A. S.; Kneipp, H.; Badizadegan, K.; Yoshizawa, N.; Boone, C.; Shafer-Peltier, K. E.; Motz, J. T.; Dasari, R. R.; Feld, M. S. *Appl. Spectrosc.*, **2002**, *56*, 150-154.
- [85] Kneipp, J.; Kneipp, H.; McLaughlin, M.; Brown, D.; Kneipp, K. *Nano Lett.*, **2006**, *6*, 2225-2231.
- [86] Kneipp, K.; Wang, Y.; Kneipp, H.; Perelman, L. T.; Itzkan, I.; Dasari, R. R.; Feld, M. S. *Phys. Rev. Lett.*, **1997**, *78*, 1667.
- [87] Nie, S.; Emory, S. R. *Science*, **1997**, *275*, 1102-1106.
- [88] Kneipp, J.; Kneipp, H.; Wittig, B.; Kneipp, K. *J. Phys. Chem. C*, **2010**.
- [89] Lyandres, O.; Shah, N. C.; Yonzon, C. R.; Walsh, J. T.; Glucksberg, M. R.; Van Duyne, R. P. *Anal. Chem.*, **2005**, *77*, 6134-6139.
- [90] Matschulat, A.; Drescher, D.; Kneipp, J. *ACS Nano*, **2010**, *4*, 3259-3269.
- [91] Feng, S. Y.; Chen, R.; Lin, J. Q.; Pan, J. J.; Chen, G. N.; Li, Y. Z.; Cheng, M.; Huang, Z. F.; Chen, J.; Zeng, H. S. *Biosens. Bioelectron.*, **2010**, *25*, 2414-2419.
- [92] Pallaoro, A.; Hoonejani, M. R.; Braun, G. B.; Meinhart, C. D.; Moskovits, M. *ACS Nano*, **2015**, *9*, 4328-4336.
- [93] Alexander, T. A. *Anal. Chem.*, **2008**, *80*, 2817-2825.
- [94] Li, S.-X.; Zeng, Q.-Y.; Li, L.-F.; Zhang, Y.-J.; Wan, M.-M.; Liu, Z.-M.; Xiong, H.-L.; Guo, Z.-Y.; Liu, S.-H. *J. Biomed. Opt.*, **2013**, *18*.
- [95] Negri, P.; Choi, J. Y.; Jones, C.; Tompkins, S. M.; Tripp, R. A.; Dluhy, R. A. *Anal. Chem.*, **2014**, *86*, 6911-6917.
- [96] Stewart, S.; Fredericks, P. M. *Spectrochim. Acta A*, **1999**, *55*, 1641-1660.
- [97] Bukowska, J.; Jackowska, K. *J. Electroanal. Chem.*, **1992**, *322*, 347-356.
- [98] Giese, B.; McNaughton, D. *J. Phys. Chem. B*, **2002**, *106*, 101-112.
- [99] Koglin, E.; Sequaris, J. M.; Valenta, P. *J. Mol. Struct.*, **1980**, *60*, 421-425.
- [100] Chen, T. T.; Kuo, C. S.; Chou, Y. C.; Liang, N. T. *Langmuir*, **1989**, *5*, 887-891.
- [101] Kneipp, K.; Kneipp, H.; Kartha, V. B.; Manoharan, R.; Deinum, G.; Itzkan, I.; Dasari, R. R.; Feld, M. S. *Physical Review E*, **1998**, *57*, R6281-R6284.
- [102] Parker, F. S., *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*. Plenum Press: New York and London, **1983**.

- [103] Webster, G. T.; De Villiers, K. A.; Egan, T. J.; Deed, S.; Tilley, L.; Tobin, M. J.; Bambery, K. R.; McNaughton, D.; Wood, B. R. *Anal. Chem.*, **2009**, *81*, 2516-2524.
- [104] Goodacre, R. *Vib. Spectrosc.*, **2003**, *32*, 33-45.
- [105] Cho, Y. T.; Su, H.; Huang, T. L.; Chen, H. C.; Wu, W. J.; Wu, P. C.; Wu, D. C.; Shiea, J. *Clin. Chim. Acta*, **2013**, *415*, 266-275.
- [106] Xu, X. Q.; Leow, C. K.; Lu, X.; Zhang, X. G.; Liu, J. S.; Wong, W. H.; Asperger, A.; Deininger, S.; Leung, H. C. E. *Proteomics*, **2004**, *4*, 3235-3245.
- [107] Tilton, R. G.; Haidacher, S. J.; LeJeune, W. S.; Zhang, X.; Zhao, Y.; Kurosky, A.; Brasier, A. R.; Denner, L. *Proteomics*, **2007**, *7*, 1729-1742.
- [108] Han, T.; Liu, H.; Yu, Z. L.; Li, J.; Wang, L.; Xiao, S. X.; Li, Y.; Yu, M. L. *J. Viral Hepat.*, **2010**, *17*, 18-23.
- [109] Rajalahti, T.; Kroksveen, A. C.; Arneberg, R.; Berven, F. S.; Vedeler, C. A.; Myhr, K. M.; Kvalheim, O. M. *J. Proteome Res.*, **2010**, *9*, 3608-3620.
- [110] Zhang, J. X.; Wu, X. Q.; Shi, L. L.; Liang, Y.; Xie, Z. S.; Yang, Y. R.; Li, Z. X.; Liu, C. H.; Yang, F. C. *Clin. Chim. Acta*, **2012**, *413*, 883-887.
- [111] Willingale, R.; Jones, D. J. L.; Lamb, J. H.; Quinn, P.; Farmer, P. B.; Ng, L. L. *Proteomics*, **2006**, *6*, 5903-5914.
- [112] Marshall, J.; Kupchak, P.; Zhu, W. M.; Yantha, J.; Vrees, T.; Furesz, S.; Jacks, K.; Smith, C.; Kireeva, I.; Zhang, R.; Takahashi, M.; Stanton, E.; Jackowski, G. *J. Proteome Res.*, **2003**, *2*, 361-372.
- [113] Agar, N. Y. R.; Malcolm, J. G.; Mohan, V.; Yang, H. W.; Johnson, M. D.; Tannenbaum, A.; Agar, J. N.; Blacks, P. M. *Anal. Chem.*, **2010**, *82*, 2621-2625.
- [114] Reyzer, M. L.; Caprioli, R. M. *Curr. Opin. Chem. Biol.*, **2007**, *11*, 29-35.
- [115] Zhang, D. H.; Tai, L. K.; Wong, L. L.; Chiu, L. L.; Sethi, S. K.; Koay, E. S. C. *Mol. Cell. Proteomics*, **2005**, *4*, 1686-1696.
- [116] Villanueva, J.; Philip, J.; Entenberg, D.; Chaparro, C. A.; Tanwar, M. K.; Holland, E. C.; Tempst, P. *Anal. Chem.*, **2004**, *76*, 1560-1570.
- [117] Villanueva, J.; Shaffer, D. R.; Philip, J.; Chaparro, C. A.; Erdjument-Bromage, H.; Olshen, A. B.; Fleisher, M.; Lilja, H.; Brogi, E.; Boyd, J.; Sanchez-Carbayo, M.; Holland, E. C.; Cordon-Cardo, C.; Scher, H. I.; Tempst, P. *J. Clin. Invest.*, **2006**, *116*, 271-284.
- [118] McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R. *Anal. Chem.*, **2005**, *77*, 6118-6124.

- [119] Fraser, P. D.; Enfissi, E. M. A.; Goodfellow, M.; Eguchi, T.; Bramley, P. M. *Plant J.*, **2007**, *49*, 552-564.
- [120] Moore, S. E. M.; Hemsley, A. R.; French, A. N.; Dudley, E.; Newton, R. P. *Protoplasma*, **2006**, *228*, 151-157.
- [121] Huecas, S.; Villalba, M.; Rodriguez, R. *J. Biol. Chem.*, **2001**, *276*, 27959-27966.
- [122] Napoli, A.; Aiello, D.; Di Donna, L.; Sajjad, A.; Perri, E.; Sindona, G. *Anal. Chem.*, **2006**, *78*, 3434-3443.
- [123] Sheoran, I. S.; Ross, A. R. S.; Olson, D. J. H.; Sawhney, V. K. *J. Exp. Bot.*, **2007**, *58*, 3525-3535.
- [124] Lauer, I.; Foetisch, K.; Kolarich, D.; Ballmer-Weber, B. K.; Conti, A.; Altmann, F.; Vieths, S.; Scheurer, S. *Biochem. J.*, **2004**, *383*, 327-334.
- [125] Lauer, I.; Alessandri, S.; Pokoj, S.; Reuter, A.; Conti, A.; Vieths, S.; Scheurer, S. *Mol. Nutr. Food Res.*, **2008**, *52*, S262-S271.
- [126] Alisi, C.; Afferni, C.; Iacovacci, P.; Barletta, B.; Tinghino, R.; Butteroni, C.; Puggioni, E. M. R.; Wilson, I. B. H.; Federico, R.; Schinina, M. E.; Ariano, R.; Di Felice, G.; Pini, C. *Allergy*, **2001**, *56*, 978-984.
- [127] Diaz-Perales, A.; Lombardero, M.; Sanchez-Monge, R.; Garcia-Selles, F. J.; Pernas, M.; Fernandez-Rivas, M.; Barber, D.; Salcedo, G. *Clin. Exp. Allergy*, **2000**, *30*, 1403-1410.
- [128] Raftery, M. J.; Saldanha, R. G.; Geczy, C. L.; Kumar, R. K. *Respir. Res.*, **2003**, *4*, 10.
- [129] Pallua, J. D.; Recheis, W.; Poeder, R.; Pfaller, K.; Pezzei, C.; Hahn, H.; Huck-Pezzei, V.; Bittner, L. K.; Schaefer, G.; Steiner, E.; Andre, G.; Hutwimmer, S.; Felber, S.; Pallua, A. K.; Pallua, A. F.; Bonn, G. K.; Huck, C. W. *Analyst*, **2012**, *137*, 1584-1595.
- [130] Welker, M. *Proteomics*, **2011**, *11*, 3143-3153.
- [131] Lasch, P.; Nattermann, H.; Erhard, M.; Stammeler, M.; Grunow, R.; Bannert, N.; Appel, B.; Naumann, D. *Anal. Chem.*, **2008**, *80*, 2026-2034.
- [132] Lasch, P.; Drevinek, M.; Nattermann, H.; Grunow, R.; Stammeler, M.; Dieckmann, R.; Schwecke, T.; Naumann, D. *Anal. Chem.*, **2010**, *82*, 8464-8475.
- [133] Ernst, M.; Silva, D. B.; Silva, R.; Monge, M.; Semir, J.; Vêncio, R. Z. N.; Lopes, N. P. *Anal. Chim. Acta*, **2015**, *859*, 46-58.
- [134] Krause, B.; Seifert, S.; Panne, U.; Kneipp, J.; Weidner, S. M. *Rapid Commun. Mass Spectrom.*, **2012**, *26*, 1032-1038.
- [135] Hassani, S.; Hanafi, M.; Qannari, E.; Kohler, A. *Chemometr. Intell. Lab.*, **2013**, *120*, 154-168.

- [136] Hassani, S.; Martens, H.; Qannari, E.; Kohler, A. *Chemometr. Intell. Lab.*, **2012**, *118*, 246-259.
- [137] Hassani, S.; Martens, H.; Qannari, E. M.; Hanafi, M.; Borge, G. I.; Kohler, A. *Chemometr. Intell. Lab.*, **2010**, *104*, 140-153.
- [138] Bras, L. P.; Bernardino, S. A.; Lopes, J. A.; Menezes, J. C. *Chemometr. Intell. Lab.*, **2005**, *75*, 91-99.
- [139] Smilde, A. K.; van der Werf, M. J.; Bijlsma, S.; van der Werff-van-der Vat, B. J. C.; Jellema, R. H. *Anal. Chem.*, **2005**, *77*, 6729-6736.
- [140] Skov, T.; Ballabio, D.; Bro, R. *Anal. Chim. Acta*, **2008**, *615*, 18-29.
- [141] Irudayaraj, J.; Xu, F.; Tewari, J. *J. Food Sci.*, **2003**, *68*, 2040-2045.
- [142] Bertelli, D.; Plessi, M.; Sabatini, A. G.; Lolli, M.; Grillenzoni, F. *Food Chem.*, **2007**, *101*, 1565-1570.
- [143] He, J.; Rodriguez-Saona, L. E.; Giusti, M. M. *J. Agric. Food Chem.*, **2007**, *55*, 4443-4452.
- [144] Soriano, A.; Perez-Juan, P. M.; Vicario, A.; Gonzalez, J. M.; Perez-Coello, M. S. *Food Chem.*, **2007**, *104*, 1295-1303.
- [145] Nicolaou, N.; Goodacre, R. *Analyst*, **2008**, *133*, 1424-1431.
- [146] Tapp, H. S.; Defernez, M.; Kemsley, E. K. *J. Agric. Food Chem.*, **2003**, *51*, 6110-6115.
- [147] Lerma-Garcia, M. J.; Ramis-Ramos, G.; Herrero-Martinez, J. M.; Simo-Alfonso, E. F. *Food Chem.*, **2010**, *118*, 78-83.
- [148] Barth, A. *Biochimica Et Biophysica Acta-Bioenergetics*, **2007**, *1767*, 1073-1101.
- [149] Belbachir, K.; Noreen, R.; Gouspillou, G.; Petibois, C. *Anal. Bioanal. Chem.*, **2009**, *395*, 829-837.
- [150] Gazi, E.; Dwyer, J.; Gardner, P.; Ghanbari-Siahkali, A.; Wade, A. P.; Miyan, J.; Lockyer, N. P.; Vickerman, J. C.; Clarke, N. W.; Shanks, J. H.; Scott, L. J.; Hart, C. A.; Brown, M. *J. Pathol.*, **2003**, *201*, 99-108.
- [151] Gajjar, K.; Trevisan, J.; Owens, G.; Keating, P. J.; Wood, N. J.; Stringfellow, H. F.; Martin-Hirsch, P. L.; Martin, F. L. *Analyst*, **2013**, *138*, 3917-3926.
- [152] Bhargava, R. *Anal. Bioanal. Chem.*, **2007**, *389*, 1155-1169.
- [153] Kohler, A.; Bertrand, D.; Martens, H.; Hannesson, K.; Kirschner, C.; Ofstad, R. *Anal. Bioanal. Chem.*, **2007**, *389*, 1143-1153.
- [154] Laurens, L. M. L.; Wolfrum, E. J. *Bioenergy Research*, **2011**, *4*, 22-35.

- [155] Chen, H.; Ferrari, C.; Angiuli, M.; Yao, J.; Raspi, C.; Bramanti, E. *Carbohydr. Polym.*, **2010**, 82, 772-778.
- [156] Kirschner, C.; Maquelin, K.; Pina, P.; Thi, N. A. N.; Choo-Smith, L. P.; Sockalingum, G. D.; Sandt, C.; Ami, D.; Orsini, F.; Doglia, S. M.; Allouch, P.; Mainfait, M.; Puppels, G. J.; Naumann, D. *J. Clin. Microbiol.*, **2001**, 39, 1763-1770.
- [157] Kim, S. W.; Ban, S. H.; Chung, H.; Cho, S.; Chung, H. J.; Choi, P. S.; Yoo, O. J.; Liu, J. R. *Plant Cell Rep.*, **2004**, 23, 246-250.
- [158] Mohlenhoff, B.; Romeo, M.; Diem, M.; Woody, B. R. *Biophys. J.*, **2005**, 88, 3635-3640.
- [159] Kohler, A.; Bocker, U.; Warringer, J.; Blomberg, A.; Omholt, S. W.; Stark, E.; Martens, H. *Appl. Spectrosc.*, **2009**, 63, 296-305.
- [160] Bambery, K. R.; Wood, B. R.; McNaughton, D. *Analyst*, **2012**, 137, 126-132.
- [161] Bassan, P.; Sachdeva, A.; Kohler, A.; Hughes, C.; Henderson, A.; Boyle, J.; Shanks, J. H.; Brown, M.; Clarke, N. W.; Gardner, P. *Analyst*, **2012**, 137, 1370-1377.
- [162] Jimenez-Hernandez, M.; Hughes, C.; Bassan, P.; Ball, F.; Brown, M. D.; Clarke, N. W.; Gardner, P. *Analyst*, **2013**, 138, 3957-3966.
- [163] Zimmermann, B.; Kohler, A. *Appl. Spectrosc.*, **2013**, 67, 892-902.
- [164] Liland, K. H.; Kohler, A.; Shapaval, V. *Chemometr. Intell. Lab.*, **2014**, 138, 41-47.
- [165] Zimmermann, B.; Bağcıoğlu, M.; Sandt, C.; Kohler, A. *Planta*, **2015**, 1-14.
- [166] Rebenstorf, K.; Candresse, T.; Dulucq, M. J.; Buttner, C.; Obermeier, C. *J. Virol.*, **2006**, 80, 2453-2462.
- [167] von Barga, S.; Grubits, E.; Jalkanen, R.; Buttner, C. *Silva Fenn.*, **2009**, 43, 727-738.
- [168] Buchhop, J.; von Barga, S.; Buttner, C. *J. Virol. Methods*, **2009**, 157, 147-154.
- [169] Schuster, C. E.; Miller, P. W. *Phytopathology*, **1933**, 23, 408-409.
- [170] Posnette, A. F.; Cropley, R. *Ann Rept East Mailing Res Sta [England]*, **1955**, 1954, 126-127.
- [171] Cooper, J. I.; Atkinson, M. A. *Forestry*, **1975**, 48, 193-&.
- [172] Jones, A. T.; Kinninmo, A.; Roberts, I. M. *J. Gen. Virol.*, **1973**, 18, 61-64.
- [173] Murrant, A. F.; Taylor, M.; Duncan, G. H.; Raschke, J. H. *J. Gen. Virol.*, **1981**, 53, 321-332.

- [174] von Bargaen, S.; Langer, J.; Robel, J.; Rumbou, A.; Buttner, C. *Virus Res.*, **2012**, *163*, 678-683.
- [175] Jones, A. T.; Mayo, M. A. *J. Gen. Virol.*, **1972**, *16*, 349-&.
- [176] Borja, M.; Sanchez, F.; Rowhani, A.; Bruening, G.; Ponz, F. *Virus Genes*, **1995**, *10*, 245-252.
- [177] Rumbou, A.; von Bargaen, S.; Buttner, C. *Eur. J. Plant Pathol.*, **2009**, *124*, 527-532.
- [178] Mink, G. I. *Annu. Rev. Phytopathol.*, **1993**, *31*, 375-402.
- [179] Werner, R.; Muhlbach, H. P.; Buttner, C. *Eur. J. For. Pathol.*, **1997**, *27*, 309-318.
- [180] Akira, S.; Uematsu, S.; Takeuchi, O. *Cell*, **2006**, *124*, 783-801.
- [181] Casu, B.; Reggiani, M. *J. Polym. Sci. Pol. Sym.*, **1964**, *7*, 171-185.

10.5 Publikationsverzeichnis

Artikel

V. Joseph; M. Gensler; **S. Seifert**; U. Gernert; J. P. Rabe; J. Kneipp, “Nanoscopic Properties and Application of Mix-and-Match Plasmonic Surfaces for Microscopic SERS”, *J. Phys. Chem. C* **2012**, *116*, 6859-6865.

B. Krause; **S. Seifert**; U. Panne; J. Kneipp; S. M. Weidner, “Matrix-assisted laser desorption/ionization mass spectrometric investigation of pollen and their classification by multivariate statistics”, *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1032-1038.

D. Drescher; I. Zeise; H. Traub; G. Guttman; **S. Seifert**; T. Büchner; N. Jakubowski; G. Schneider; J. Kneipp, “In situ characterization of SiO₂ nanoparticle biointeractions using Bright Silica”, *Advanced Functional Materials* **2014**, *24*, 3765-3775.

S. Seifert, S. M. Weidner, U. Panne, J. Kneipp, “Taxonomic relationships of pollens from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry data using multivariate statistics“, *Rapid Commun. Mass Spectrom.* **2015**, *29*, 1145-1154.

S. Seifert, V. Merk, J. Kneipp, “Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods“, *J. Biophotonics* **2016**, *9*, 181-189.

M. Joester, **S. Seifert**, F. Emmerling, J. Kneipp, „Physiological influence of silica on germinating pollen as shown by Raman spectroscopy“, (zur Publikation eingereicht).

S. Seifert, B. Zimmermann, M. Bağcıoğlu, A. Kohler, J. Kneipp, „Classification of single pollen grains with FTIR microspectroscopy“, (in Vorbereitung).

Vorträge

S. Seifert (Vortragender), B. Krause; U. Panne; S. Weidner; J. Kneipp “Classification and characterisation of pollen grains with Raman microspectroscopy and mass spectrometry methods”, Anakon in Essen, März 2013.

S. Seifert (Vortragender) “Characterization and classification of pollen grain samples using SERS and MALDI-TOF MS in combination with multivariate statistics”, 8. Interdisziplinäres Doktorandenseminar des GDCH Arbeitskreises Prozessanalytik in Berlin, Februar 2014.

Poster

S. Seifert; J. Kneipp “Surface Enhanced Raman Scattering on aqueous pollen extracts”, 8. RKI Workshop FT-IR Spectroscopy in Microbiological and Medical Diagnostics in Berlin, Oktober 2011.

S. Seifert; J. Kneipp “Classification of aqueous pollen extracts via Surface Enhanced Raman Scattering”, 2. Berliner Chemie Symposium in Berlin, April 2012.

S. Seifert; V. Joseph; J. Kneipp “A new approach in pollen classification: Surface Enhanced Raman Scattering on aqueous extracts”, Tag der Chemie in Berlin, Juni 2012.

S. M. Weidner; B. Enthaler; **S. Seifert**, B. C. Krause “MALDI Imaging MS of intact pollen grains and pollen mixtures”, ASMS, Juni 2013.

S. Seifert; U. Panne; J. Kneipp “Characterization and classification of aqueous pollen extracts using SERS and multivariate statistics”, 9. RKI Workshop FT-IR Spectroscopy in Microbiological and Medical Diagnostics in Berlin, Oktober 2013.

Büchner, T.; Drescher, D., **Seifert, S.**; Kneipp, J. “SERS investigations of isolated nuclei”, 9. RKI Workshop FT-IR Spectroscopy in Microbiological and Medical Diagnostics in Berlin, Oktober 2013.

S. Seifert; U. Panne; J. Kneipp „Investigation of pollen grain extracts with SERS and multivariate statistics”, ICORS in Jena, August 2014.

I. Zeise, D. Drescher, H. Traub, P. Guttman, **S. Seifert**, T. Büchner, N. Jakubowski, G. Schneider, J. Kneipp “BrightSilica – A new plasmonic tool for bioanalytical applications”, ICORS in Jena, August 2014.

M. Joester, **S. Seifert**, F. Emmerling, J. Kneipp „Raman microspectroscopy and scattering methods to investigate silica in plants”, ICORS in Jena, August 2014.

S. Seifert; S. Weidner; U. Panne; J. Kneipp “MALDI-TOF MS in combination with multivariate statistics to delineate taxonomic relationships of pollens”, Anakon in Graz, März 2015.

M. Bağcıoğlu¹, B. Zimmermann, **S. Seifert**, J. Kneipp, A. Kohler „High-throughput Fourier transform infrared spectroscopic approach for identification and biochemical characterization of pollen”, ESCBM in Bochum, September 2015.

S. Seifert; V. Merk; J. Kneipp, “Classification and identification of aqueous pollen extracts using SERS and artificial neural networks (ANN)”, 10. RKI Workshop FT-IR Spectroscopy in Microbiological and Medical Diagnostics in Berlin, Oktober 2015.

M. Joester, **S. Seifert**, J. Kneipp, F. Emmerling, „Raman micro-spectroscopy and multivariate statistics to study the influence of silica on germinated pollen grains” RKI Workshop FT-IR Spectroscopy in Microbiological and Medical Diagnostics in Berlin, Oktober 2015.